THE CONSTRUCTION AND ROLE OF NON-COVALENT BENCHMARKS IN COMPUTATIONAL CHEMISTRY

A Thesis
Presented to
The Academic Faculty

by

Michael S. Marshall

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
[August 2012]

THE CONSTRUCTION AND ROLE OF NON-COVALENT BENCHMARKS IN COMPUTATIONAL CHEMISTRY

Approved by:

Professor Jean-Luc Brédas,
Committee Chair
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor C. David Sherrill, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Charles Liotta
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Angelo Bongiorno
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Steve Harvey
School of Biology
Georgia Institute of Technology

Date Approved: [August 2012]

This thesis is dedicated to my parents,

Edward M. Marshall and Aldeana J. Marshall,

without whose love and support,

none of this would have been possible.

ACKNOWLEDGEMENTS

I would like to thank my thesis adviser Dr. C. David Sherrill for teaching me the skills needed to do well in the area of computational chemistry. His teaching style of learning by just tackling the problem head-on and picking up the knowledge as we went has been an exciting experience that not only works, but keeps our group as one of the best in the country for anything involving non-bonded interactions. The flexibility of working on projects that pique my interest (as opposed to only working towards the grant that pays the bills) has made graduate school one of the most exciting times of my life. There is a great feeling of accomplishment seeing one of your original ideas from thought to published paper that can never be replaced. For creating one of the best work environments a graduate student could ever ask for, I thank Dr. Sherrill.

I would not be in the field of computational chemistry if it were not for my undergraduate adviser, Dr. Tracy P. Hamilton. My sophomore year of undergraduate studies, I was introduced to him through a good friend, Dr. Anthony Lang. Dr. Hamilton handed me a copy of "Modern Quantum Chemistry" by Szabo and Ostlund, a couple of files of one- and two-electron integrals, and we proceeded to work through every problem through chapter three and write a Hartree–Fock code. With this knowledge, we proceeded to work on many projects ranging from high energy materials (nitro-substituted prismanes) to chemo-preventive cures for breast cancer (retinoic acid derivatives). Dr. Hamilton is a brilliant teacher, and I had the luxury of being taught by him one-on-one for a number of years. When I took his quantum class, we would always go back to the lab and discuss it in depth. I also thank him for teaching me what good beer should taste like. When it was time to graduate, he is the one

that recommended that I work for Dr. Sherrill. He was more than an adviser, he was a good friend.

I thank the University of Alabama at Birmingham Honors program, with particular love for Dr. Michael Sloane and Dr. Ada Long. Barely having the scores to get into college from highschool, I did not think I had a shot at getting into their program, but Dr. Long believed in me more than I even believed in myself at the time. The only reason she gave me for accepting me into the program with my abysmal scores was, "I see something great in your eyes," and I worked hard to try to live up to whatever she saw in me. Later, when I told them I had to take some time off to save money to pay for college, they found me the money to stay through a variety of scholarships, which paid for the rest of my time at UAB. Without them, I probably would not have even finished my undergraduate degree. Thank you for creating one the greatest honors programs in the country, and thank you for caring about me the way you did.

I would like to thank past and present Sherrill group members, many of which are as close as family by now. I would like to especially thank Dr. Ed Hohenstein and Dr. John Sears for the always being there to bounce ideas around with and we sure did have a lot of bad ones–remember triangular basis sets or Hartree based MP2? You were a both great friends and even better scientists. Thank you for always being there for me. I would like to thank Dr. Lori Burns for the many projects we have worked on together and for always editing away my Alabama grammar. We recently wrote a paper together where one reviewer said, "this paper should be accepted for its writing alone". I forwarded the comment to Dr. Burns because I knew it was not for my contribution. I would like to thank Dr. Sahan Thanthiriwatte, Dr. Tait Takatani, Yue Geng, Dr. Ashley Ringer, Stephen Arnstein, Dr. Eugene DePrince, Matthew Kennedy, Rob Parrish, Trent Parker, Vernon Crowell, and Dr. Massimo Malagoli for all the great conversations we have had at GaTech.

Finally, I would like to thank my parents, Edward M. Marshall and Aldeana J. Marshall. They have always been there for me, unconditionally. At a young age, my dad taught me math and science well ahead of where we were in school, and my mom taught me how to enjoy life and bring out the artsy side of me. They were never the typical parents that tell you what your next best step in life is and force you down their path, instead they just encouraged me to be happy and see where that takes me. I may never win raffles or the lottery, but I won the birth lottery being given to them. With all the love in my heart, thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The focus of this thesis is on improving the description and understanding of non-bonded interactions. Non-bonding interactions can range in strength from very weak dispersion-bound complexes such as rare gas dimers to very strong electrostatically bound salt bridges such as lysine interacting with aspartic acid. Of particular interest to chemistry is the interaction energy of two monomers, defined as

$$E_{IE} = E_{AB} - E_A - E_B, \tag{1}$$

where subscript AB indicates the dimer and subscripts A and B indicate the monomers A and B, respectively. This quantity is the energy difference between monomers at a finite separation compared to isolated monomers (infinite separation). In this definition, a negative interaction energy means a favorable interaction. To better understand interactions such as hydrogen bonding, C-H/π, and π-π interactions, prototypical complexes are commonly examined. For example, the benzene dimer is often the model system for studying π-π interactions. It is to these prototype complexes that very high level computations are then applied. The current state of the art method available for molecular systems of less than $\sim$40 atoms is coupled-cluster theory with iterative single and double excitations and perturbative triple excitations [CCSD(T)], which will be detailed in chapter 1. Using CCSD(T) in conjunction with large basis sets often results in good agreement with experimental results. It is with these very high level computations that we can then start to learn exactly what physics is being captured properly with more routinely used low scaling methods such as force-fields, semi-empirical methods, and density functional theory. As computational power has increased, so has the number of accessible prototype systems that we can investigate,

and this gives rise to benchmark databases, which are collections of accurate benchmarks. The most commonly used benchmark database in the area of non-bonded interactions is the S22 database,[103] which is comprised of 22 non-bonded interactions ranging from water dimer (6 atoms) to adenine·thymine complexes (30 atoms).

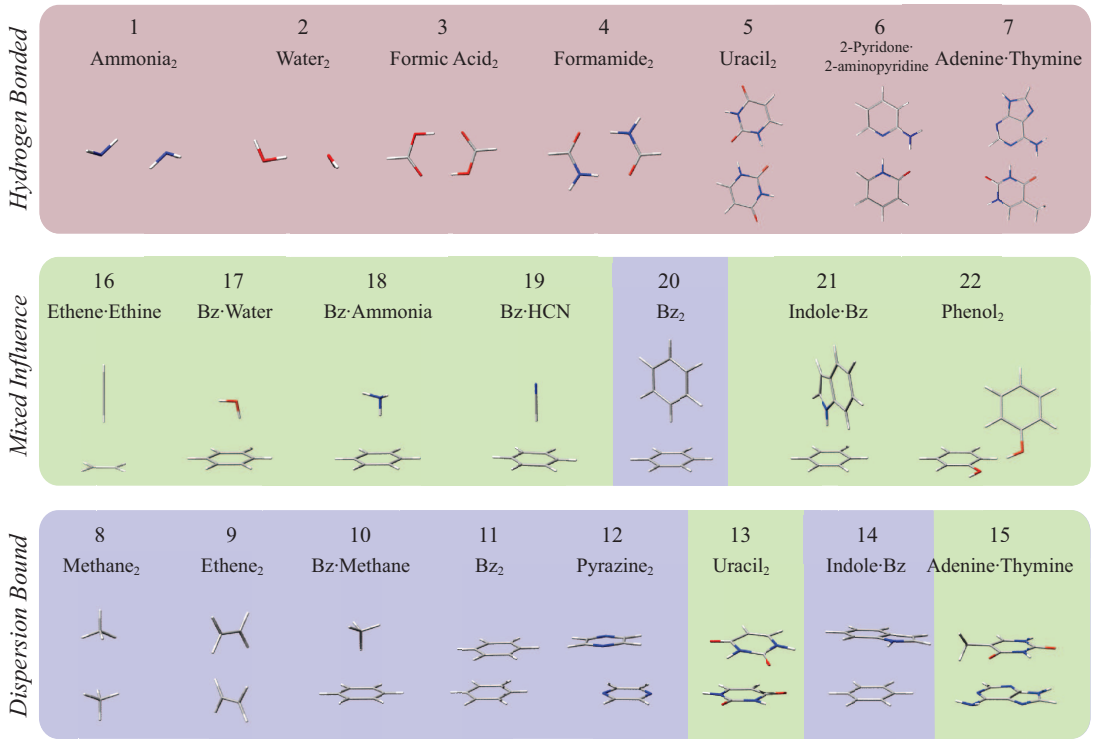

Figure 1: The S22 benchmark database by binding motif. Red, blue, and green represent hydrogen-bound, dispersion, and mixed-influence interactions, respectively. Some complexes have had their binding motif redefined with symmetry adapted perturbation theory, illustrated by complexes being in the wrong row. Reproduced with permission from Lori A. Burns.

Shown in Figure 1, the S22 database attempts to capture the important non-bonded interactions commonly found in biological systems. Some of these complexes are so large that even CCSD(T) cannot be used without approximations. The most common approximation when trying to achieve very accurate interaction energies for systems as large as adenine·thymine complexes is to use focal-point

analysis.[48, 38] Focal-point analysis allows one to estimate CCSD(T) in a large basis [est. CCSD(T)/large]:

$$E_{CCSD(T)}^{large} \approx E_{MP2}^{large} + \delta_{MP2}^{CCSD(T)}, \tag{2}$$

$$\delta_{MP2}^{CCSD(T)} = E_{CCSD(T)}^{small} - E_{MP2}^{small}, \tag{3}$$

where the subscripts and superscripts on energies refer to the method and basis set, respectively. The "large" basis set is often a complete basis set (CBS) extrapolation, whereas the "small" basis set used for $\delta_{MP2}^{CCSD(T)}$ is usually a single basis set, though it may also be a basis set extrapolation.[193] As has been noted,[199, 101, 82, 198, 155] the focal-point approach works well because even though the convergence rates of CCSD(T) and MP2 correlation energies are slow with respect to basis set size, the rate of convergence of the difference [CCSD(T)–MP2] is much faster.[100] Thus, a much smaller basis set may be used for the $\delta_{MP2}^{CCSD(T)}$ correction than for the underlying SCF and MP2 computations. This approach has been applied to obtain several non-covalent interaction benchmark energies.[221, 199, 103, 50]

To better understand the errors remaining in our benchmark computations, each component in the focal-point analysis was looked at individually. A complete understanding of the convergence of the MP2 component, as well as common approximations made to MP2 were investigated (Chapter 4). We demonstrate that with a careful choice of approximations, MP2-quality results can be computationally affordable for systems with a few dozen atoms or larger without introducing significant error. Density fitting (explained in detail in chapter 1) reduces the time to compute the MP2 correlation energy, dual-basis techniques abate the cost of the underlying SCF, and heavy-augmented functions (using aug-cc-pVNZ on heavy atoms, and cc-pVNZ on hydrogen) speed up both parts of the computation relative to the fully

augmented basis sets. Except for comparisons using the smaller heavy-aug-cc-pVDZ basis set, all of these approximations show significant speed-up while never incurring a RMSE greater than 0.045 kcal mol^{-1} for the S22 test cases. We also demonstrate that all of these approximations do indeed combine very efficiently.

The more complicated contribution to focal-point methods is $\delta_{MP2}^{CCSD(T)}$, the "coupled-cluster correction." It is very common to find literature that claims "benchmark" quality computations while using $\delta_{MP2}^{CCSD(T)}$ corrections with very small basis sets, so a very detailed study of basis set effects employing common polarized double-ζ sets used in "benchmark" quality results was performed (Chapter 3). We show that although double-ζ quality bases generally yield adequate estimates of $\delta_{MP2}^{CCSD(T)}$, especially if diffuse functions are included (or if the d exponent is made more diffuse), the errors of 0.10-0.39 kcal mol^{-1} for the S22 molecules are too large for the resulting CCSD(T)/CBS values to be of true benchmark quality, given that several approximate methods are now capable of reproducing benchmark interaction energies within a few tenths of one kcal mol^{-1}. Hence, the remaining errors in $\delta_{MP2}^{CCSD(T)}$ should be taken into consideration when comparing new methods against benchmark sets that utilize $\delta_{MP2}^{CCSD(T)}$ corrections with such basis sets. By examining the coupled-cluster correction in progressively higher angular momentum basis sets, up to aug-cc-pV6Z, a characteristic turning point was found, after which the quantity converges monotonically and before which CBS extrapolations are unreliable. Particularly, CBS extrapolated $\delta_{MP2}^{CCSD(T)}$ corrections should not be used for hydrogen bonded complexes when employing aDZ and aTZ basis sets for the extrapolation. We demonstrate that simply using the largest single basis set affordable is often the best choice.

Having computed reliable benchmarks and investigated the remaining intrinsic errors, these types of benchmarks can be used to learn about interesting non-bonded complexes (such as side-on cation-π in chapter 2) or fit existing methods against these accurate values to improve accuracy (such as DW-CCSD(T**)-F12 in chapter

6). The latter reduces the accessible double-ζ mean errors for S22 by a factor of two and the max error by a factor of 3.

CHAPTER I

INTRODUCTION

1.1 Fundamentals of Electronic Structure Theory

The purpose of this section is to provide a basic introduction into electronic structure theory as it pertains to this thesis. The notation used throughout this chapter is that adopted by Szabo and Ostlund.[209]

1.1.1 Schrödinger Equation

With a few exceptions, the field of electronic structure theory aims at solving the time-independent, non-relativistic Schrödinger equation shown as equation 4,

$$\hat{\mathcal{H}}\Psi = \mathcal{E}\Psi, \tag{4}$$

where $\hat{\mathcal{H}}$ represents the Hamiltonian operator and Ψ is the molecular wavefunction. For a system with N electrons and M atoms, the Hamiltonian operator in atomic units is defined as

$$\hat{\mathcal{H}} = \overbrace{-\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2}^{(a)} \overbrace{-\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2}^{(b)} \overbrace{-\sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}}}^{(c)} + \overbrace{\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}}}^{(d)} + \overbrace{\sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}}^{(e)}, \tag{5}$$

where ∇_i^2 and ∇_A^2 are Lapacian operators that involve differentiation with respect to the coordinates of the ith electron and Ath nucleus, respectively, M_A represents the mass of the Ath nucleus, Z_A represent the charge of the Ath nucleus, r_{ij} represents the distance between the ith and the jth electron, R_{AB} represents the distance between the Ath nucleus and the Bth nucleus, and R_{iA} represents the distance between the ith electron and the Ath nucleus. Each summation has a physical interpretation as follows (a) the total kinetic energy of all electrons, (b) the total kinetic energy of all

nuclei, (c) the total coulombic attraction between all electrons and nuclei, (d) the total electron-electron repulsion, and (e) the total nuclear-nuclear repulsion energy. From here, a series of approximations will be applied to result in tractable equations for reasonable sized systems. The first approximation is in how relativistic effects are handled. For first- and second-row chemistry, relativistic effects are of little concern and are very often ignored, as large errors only emerge for 4th row and beyond, for which correcting/compensating pseudo-potentials are becoming very reliable.

An important approximation in making the Schrödinger equation solvable is the Born-Oppenheimer approximation. Since the nuclei are significantly more massive ($1.674*10^{-27}$ kilograms for the lightest atom, hydrogen) than an electron ($9.109*10^{-31}$ kilograms), one can approximate the correlated movements of nuclei and electrons with the movement of the electrons in a field of fixed nuclei. When this approximation is employed, the (b) term in equation 5 will be zero since the nuclei are no longer allowed to be moving and the (e) term becomes a constant formally known as the nuclear repulsion energy. Since this value is a constant it is simply added to the eigenvalues and has no effect on the operator eigenfunctions. The remaining terms (a,c, and d) form what is known as the electronic Hamiltonian (Equation 6) that will fully describe the electrons moving in the presence of fixed nuclei.

$$\hat{\mathcal{H}}_{elec} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} \tag{6}$$

While equation 6 looks much simpler, it is still not exactly analytically solvable for any molecular system with more than one electron. From here, electronic structure theorists have to start making more severe approximations on how to handle this many-body problem. To achieve tractability, wavefunctions must be viewed as a product of one-electron functions,

$$\Psi^{HP}(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N) = \chi_i(\mathbf{x}_1)\chi_j(\mathbf{x}_2)\chi_k(\mathbf{x}_3)\cdots\chi_k(\mathbf{x}_N). \tag{7}$$

To enforce anti-symmetry in the resulting wavefunction, a single Slater determinant

is used.

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_i(\mathbf{x}_1) & \chi_j(\mathbf{x}_1) & \dots & \chi_k(\mathbf{x}_1) \\ \chi_i(\mathbf{x}_2) & \chi_j(\mathbf{x}_2) & \dots & \chi_k(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ \chi_i(\mathbf{x}_N) & \chi_j(\mathbf{x}_N) & \dots & \chi_k(\mathbf{x}_N) \end{vmatrix} \tag{8}$$

where $\frac{1}{\sqrt{N!}}$ is the normalization factor and $\chi(\mathbf{x})$ represents the one-electron spin orbitals. We do not yet know what the spin orbitals χ_i should look like and we will have to construct them using the linear combination of atomic orbitals to form molecular orbitals (LCAO-MO) approach,

$$\chi_i = \sum_{\mu} C_{\mu}^{i} \phi_{\mu}, \tag{9}$$

where C_{μ}^{i} are expansion coefficients and ϕ_{μ} are atomic orbitals constructed from Gaussian-type orbitals (GTO),

$$\phi_{\mu}^{GTO}(r) = N x^l y^m z^n e^{-\alpha r^2}, \tag{10}$$

where l, m, and n are integers used to designate the angular momentum of the orbitals. The use of these basis functions allows one to cast everything into matrix formalism as will be demonstrated in the next chapter on Hartree–Fock Theory.

1.1.2 Hartree–Fock Theory

The central approximation to electronic structure theory is the Hartree–Fock (HF) approximation. Within this approximation, each electron can only interact with the average field exhibited by the other electrons. This reduces the last term of equation 6 from a two-electron operator to a one-electron operator. This leaves one with the Fock-operator,

$$\hat{f}(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \nu^{HF}(i), \tag{11}$$

and equations to solve of the form

$$\hat{f}(i)\chi_j(\mathbf{x}_i) = \varepsilon_j \chi_j(\mathbf{x}_i). \tag{12}$$

So at the heart of Hartree–Fock theory is the simplification of the many-body problem by replacing the two-electron part of the Hamiltonian with an effective average potential, $\nu^{HF}(i)$.

Using the above definitions, we can show that the HF energy in terms of one- and two-electron integrals becomes

$$E_{HF} = \sum_i \langle i|h|i\rangle + \frac{1}{2}\sum_{ij}[ii|jj] - [ij|ji], \tag{13}$$

where

$$\langle i|h|j\rangle = \int d\mathbf{x}_1 \chi_i^*(\mathbf{x}_1) h(\mathbf{r}_1) \chi_j(\mathbf{x}_1), \tag{14}$$

and

$$[ij|kl] = \int d\mathbf{x}_1 d\mathbf{x}_2 \chi_i^*(\mathbf{x}_1)\chi_j(\mathbf{x}_1)\frac{1}{r_{12}}\chi_k^*(\mathbf{x}_2)\chi_l(\mathbf{x}_2). \tag{15}$$

Equations 13 through 15 leave us with a path forward to iteratively solve for the proper coefficients of equation 9. The iterative procedure as well as the starting parameters are described at great length in Szabo and Ostlund.[209]

1.2 Correlated Methods

1.2.1 Møller–Plesset Perturbation Theory

Full reviews of Møller–Plesset Perturbation Theory can be found in references [15, 16, 17, 125]. This section can also be found as a presentation by me from 2010 at http://vergil.chemistry.gatech.edu/video/mbpt.html. Improving on Hartree–Fock requires a better description of the electron-electron interactions, which has only been handled in an average way up until this point. The simplest of the methods that include *dynamic electron correlation* is Møller–Plesset Theory. In this approximation, we introduce a small perturbation to the Hamiltonian as,

$$\hat{\mathcal{H}}\Psi = (\hat{\mathcal{H}}_0 + \mathcal{V})\Psi = \mathcal{E}\Psi. \tag{16}$$

To be able to systematically improve the Hamiltonian operator, we would like to be able to switch between the unperturbed state and the fully perturbed state. This is

accomplished with a single parameter λ,

$$\hat{\mathcal{H}} = (\hat{\mathcal{H}}_0 + \lambda\mathcal{V}). \tag{17}$$

Expanding the exact eigenvalues and eigenfunctions in a Taylor series results in

$$|\Psi_i\rangle = |\Psi_i^{(0)}\rangle + \lambda|\Psi_i^{(1)}\rangle + \lambda^2|\Psi_i^{(2)}\rangle + \cdots, \tag{18}$$

and

$$\mathcal{E}_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \cdots, \tag{19}$$

where $E_i^{(n)}$ and $\Psi_i^{(n)}$ are the nth-order energy and wavefunction, respectively. Substituting equations 18 and 19 back into 17 yields

$$\begin{aligned}(\hat{\mathcal{H}}_0 + \lambda\mathcal{V})(|\Psi_i^{(0)}\rangle + \lambda|\Psi_i^{(1)}\rangle + \lambda^2|\Psi_i^{(2)}\rangle + \cdots) = \\ (E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \cdots)(|\Psi_i^{(0)}\rangle + \lambda|\Psi_i^{(1)}\rangle + \lambda^2|\Psi_i^{(2)}\rangle + \cdots)\end{aligned} \tag{20}$$

After expanding equation 20, one can collect equal powers of λ because equation 20 must hold for all $\lambda \in [0, 1]$. This results in a collection of equations formally known as nth order Schrodinger equations,

$$\begin{aligned}
\lambda^0 &: \quad \hat{\mathcal{H}}_0|\Psi_i^{(0)}\rangle = E_i^{(0)}|\Psi_i^{(0)}\rangle \\
\lambda^1 &: \quad \hat{\mathcal{H}}_0|\Psi_i^{(1)}\rangle + \mathcal{V}|\Psi_i^{(0)}\rangle = E_i^{(0)}|\Psi_i^{(1)}\rangle + E_i^{(1)}|\Psi_i^{(0)}\rangle \\
\lambda^2 &: \quad \hat{\mathcal{H}}_0|\Psi_i^{(2)}\rangle + \mathcal{V}|\Psi_i^{(1)}\rangle = E_i^{(0)}|\Psi_i^{(2)}\rangle + E_i^{(1)}|\Psi_i^{(1)}\rangle + E_i^{(2)}|\Psi_i^{(0)}\rangle \\
\lambda^k &: \hat{\mathcal{H}}_0|\Psi_i^{(n)}\rangle + \mathcal{V}|\Psi_i^{(n-1)}\rangle = \sum_{n=0}^{k} E_i^{(n)}|\Psi_n^{(k-n)}\rangle
\end{aligned}$$

This treatment is typically truncated at the 2nd order, since higher order corrections are often at least as computationally demanding as more accurate correlated methods that we will examine in the following chapters. We are now left with a 2nd order correction to the Hartree–Fock reference as,

$$E_i^{(2)} = \langle\Psi_i^{(0)}|\mathcal{V}|\Psi_i^{(1)}\rangle, \tag{21}$$

and the first-order perturbed wavefunction which is constructed as a linear combination of unperturbed wavefunctions,

$$|\Psi_i^{(1)}\rangle = \sum_{\mu \neq i} \frac{\langle \Psi_\mu^{(0)} | \mathcal{V} | \Psi_i^{(0)} \rangle}{E_i^{(0)} - E_\mu^{(0)}} |\Psi_\mu^{(0)}\rangle, \tag{22}$$

and we define the perturbation as,

$$\mathcal{V} = \sum_{i<j} r_{ij}^{-1} - \mathcal{V}^{HF}. \tag{23}$$

Using Chemist's notation, we write the two-electron integrals as,

$$[ij|kl] = \int d\mathbf{x}_1 d\mathbf{x}_2 \chi_i^*(\mathbf{x}_1)\chi_j(\mathbf{x}_1)\frac{1}{r_{12}}\chi_k^*(\mathbf{x}_2)\chi_l(\mathbf{x}_2). \tag{24}$$

Using this notation, the second-order correction to the energy becomes

$$E_o^{(2)} = \frac{1}{4}\sum_{ijab} \frac{|[ai|bj] - [aj|bi]|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}, \tag{25}$$

where electrons in molecular orbitals i and j are being excited into molecular orbitals a and b. The numerator contains only two-electron integrals and the denominator is the energy difference of the occupied orbitals and the orbitals the electrons are being excited into. This equation is also more commonly represented in physicist's notation,

$$E_o^{(2)} = \frac{1}{4}\sum_{ijab} \frac{|\langle ab|ij\rangle - \langle ab|ji\rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}. \tag{26}$$

The total energy of MP2 is now the sum of the energy corrections through second-order,

$$E_{MP2} = \overbrace{E^{(0)} + E^{(1)}}^{Hartree-Fock} + E^{(2)}. \tag{27}$$

Now that electron-electron interactions are being described in more than an average way, binding types like van der Waals interactions, which are dispersion bound, can be investigated. MP2 is also routinely used for accurate geometry optimizations because it can now handle longer distance 1,3- and 1,4-interactions. While perturbation theory does provide a better description of molecular systems than Hartree–Fock, it

is not without its own defects. Primarily, it has been shown that including higher order corrections does not always provide better accuracy. Near orbital degeneracies, perturbation theory to diverges to infinity quickly, as can be seen in the denominator of equation 26 when $\epsilon_a + \epsilon_b$ approaches $\epsilon_i - \epsilon_j$. The higher the order of the perturbation theory, the faster the energy will tend to infinity artificially because of these orbital degeneracies. MP2 also suffers from severely overbinding dispersion bound complexes, especially π-π interactions.

1.2.2 Coupled-Cluster Theory

For a more complete review of Coupled-Cluster (CC) theory, the reader is directed to the following references.[170, 179, 172, 189] Coupled-cluster theory is among the most robust levels of theory that can describe dynamic electron correlation. It achieves this by using an exponential ansatz typically acting on the Hartree–Fock reference $|\Phi_0\rangle$, though not restricted to this reference,

$$|\Psi_{CC}\rangle = e^{\hat{\mathbf{T}}}|\Phi_0\rangle, \tag{28}$$

where $\hat{T}$ is the excitation operator,

$$\hat{\mathbf{T}} = \hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_2 + \hat{\mathbf{T}}_3 \cdots + \hat{\mathbf{T}}_N. \tag{29}$$

With this definition, the $\hat{\mathbf{T}}_N$ operator acts on the reference wavefunction to form excited determinants multiplied by a coefficient. For example, the $\hat{\mathbf{T}}_1$ operator will act on the reference wavefunction to form singly-excited determinants Φ_i^a multiplied by the coefficients t_i^a, which will be solved for iteratively. Analogously, the $\hat{\mathbf{T}}_2$ operator will act on the reference wavefunction to form doubly-excited determinants Φ_{ij}^{ab} with coefficients t_{ij}^{ab}. These two operators are defined as,

$$\hat{\mathbf{T}}_1\Phi_0 = \sum_{i}^{occ}\sum_{a}^{vir} t_i^a \Phi_i^a \tag{30}$$

and

$$\hat{\mathbf{T}}_2\Phi_0 = \sum_{i<j}^{occ}\sum_{a<b}^{vir} t_{ij}^{ab} \Phi_{ij}^{ab}. \tag{31}$$

Substituting equation 29 into 28 and expanding the exponential into a Taylor series results in,

$$e^{\hat{\mathbf{T}}} = \mathbf{1} + \hat{\mathbf{T}}_1 + (\hat{\mathbf{T}}_2 + \frac{1}{2}\hat{\mathbf{T}}_1^2) + (\hat{\mathbf{T}}_3 + \hat{\mathbf{T}}_2\hat{\mathbf{T}}_1 + \frac{1}{6}\hat{\mathbf{T}}_1^3) + \cdots , \tag{32}$$

where we have already grouped the operators by order of excitation. In coupled cluster theory, there are two types of interacting excitations present, connected and disconnected. An example of a connected double excitation is $\hat{\mathbf{T}}_2$, as opposed to the disconnected double excitation of $\hat{\mathbf{T}}_1^2$. The disconnected double excitation has two singly excited determinants that are non-interacting. Even truncating coupled-cluster theory to only include single and double excitations (CCSD) will include these disconnected, higher-order excitations, which contribute to make coupled-cluster theory a very accurate method for its cost.

Very often, truncating coupled-cluster theory at double excitations can significantly underestimate the energy of weakly bound complexes, but including iterative triple excitations is too computationally demanding. For these cases, perturbative triples can be included which is commonly abbreviated, CCSD(T). This method is commonly referred to as the "gold-standard" of computational chemistry and will set the bar to by which most other methods will be compared.

1.2.3 Symmetry-Adapted Perturbation Theory

Full reviews of Symmetry-Adapted Perturbation Theory (SAPT) can be found in references.[98, 236, 97, 164, 77, 88, 89, 90] Within the supermolecular *ab initio* framework, the interaction energy of two molecules A and B is defined using equation 33:

$$\Delta\tilde{E}_{int} = \tilde{E}_{AB} - \tilde{E}_A - \tilde{E}_B \tag{33}$$

Where the $\tilde{E}_{AB}$, $\tilde{E}_A$, and $\tilde{E}_B$ are approximations to the true ground-state energies E_{AB}, E_A, and E_B, which represent the energy of the dimer, monomer A, and monomer

B respectively. The errors of $E_i - \tilde{E}_i$ are always much larger than the interaction energy itself. $\tilde{E}_{int}$ can only be a good approximation to E_{int} (true interaction energy) if a fortuitous cancellation of errors occur. This very large difference in scale is often demonstrated with the example of trying to weigh a person by weighing him and an airplane and subtracting the weight of the airplane. As grateful as people are to have bathroom scales to weigh people, quantum chemists are grateful for what is referred to as symmetry-adapted perturbation theory (SAPT)[98, 237] to compute interaction energies.

With SAPT, one can compute the interaction energy directly without having to first calculate the total energy of any component, thus avoiding subtracting very large quantities. SAPT approximates the interaction energy as the sum of several physically motivated contributions:

$$E_{interaction}^{SAPT} = E_{elst}^{(1)} + E_{exch}^{(1)} + E_{ind}^{(2)} + E_{exch-ind}^{(2)} + E_{disp}^{(2)} + E_{exch-disp}^{(2)} \tag{34}$$

The superscripts refer to the order of the correction. As done in our previous studies and by others,[196, 86] we collect energy components as follows:

$$E_{ind}^{(2)} = E_{ind}^{(2)} + E_{exch-ind}^{(2)} \tag{35}$$

$$E_{disp}^{(2)} = E_{disp}^{(2)} + E_{exch-disp}^{(2)} \tag{36}$$

which results in a total interaction energy of:

$$E_{SAPT} = E_{exch}^{(1)} + E_{elst}^{(1)} + E_{disp}^{(2)} + E_{ind}^{(2)} \tag{37}$$

The following sections will attempt to explain each term found in SAPT in a more qualitative sense.

1.2.3.1 Electrostatics:

$$Elst^{(1)} = \sum_{AB} \Big[4\,(AA|BB) + 2\,(A|\nu_B|A) + 2\,(B|\nu_A|B) \Big] + V_0 \tag{38}$$

$(AA|BB)$ is a Coulomb integral between the charge distribution on monomer A and monomer B. The constant of four comes from it being a two-electron term and the respective spin states ($\alpha\alpha$, $\alpha\beta$,$\beta\alpha$,$\beta\beta$).

$(A|\nu_B|A)$ are one-electron integrals that represent the charge distribution on monomer A interacting with the nuclei of monomer B. The factor of two comes from it being a one-electron term and the respective spin states (α,β).

$(B|\nu_A|B)$ are one-electron integrals that represent the charge distribution on monomer B interacting with the nuclei of monomer A. The factor of two comes from it being a one-electron term and the respective spin states (α,β).

V_0 is the inter-monomer nuclear repulsion energy.

Note that since these equations work from computed charge distributions and not atom centered charges, we capture effects such as charge interpenetration.

1.2.3.2 Exchange:

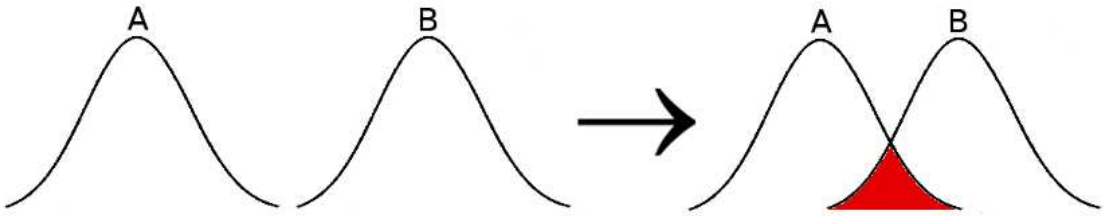

Figure 2: Overlap of two occupied orbitals from two different monomers

Exchange is by far the most complicated term in SAPT to write out (thirteen terms), but it boils down to two components (one attractive and one repulsive). The attractive term of exchange is captured in the $(AB|AB)$ term. This term can be physically thought of as a stabilizing interaction from an electron on monomer A tunneling to an orbital on monomer B (or vice versa). The dominant repulsive term arises from the Pauli exclusion principle shown graphically in figure 2. Figure 2 illustrates that at infinite separation, electrons on each monomer can distribute

themselves freely around the monomer to minimize the energy (resulting in a normal distribution of electron density). As the two monomers are brought closer to one another, occupied orbitals from monomer A start to have a non-negligible overlap with occupied orbitals on monomer B. It is in this overlap region that electrons from monomer A and monomer B share a common area of electron density. The Pauli exclusion principle restricts any two electrons of the same spin from occupying the same point in space (the wavefunction would no longer be anti-symmetric if they could). The energy it takes to redistribute the electron density (move some of the density out of the overlapped region) such that it obeys they Pauli exclusion principle in this overlapped region is the exchange-repulsion energy.

1.2.3.3 Induction:

$$Ind_{A\leftarrow B} = 2\sum_{AR}\Big[C_{AR}\,(A|\omega_B|R)\Big] \tag{39}$$

$$Ind_{B\leftarrow A} = 2\sum_{BR}\Big[C_{BR}\,(B|\omega_A|R)\Big] \tag{40}$$

$$Ind(total) = Ind_{A\leftarrow B} + Ind_{B\leftarrow A} \tag{41}$$

Induction can be thought of physically as monomer A relaxing its electron density in the presence of monomer B's electrostatic potential. The $(A|\omega_B|R)$ integrals have two competing terms: $(A|\nu_B|R)$ and $\sum_B^{occ}\Big[(AR|BB)\Big]$, which describe the interactions of the electrons on monomer A with the electrostatic potential of monomer B. The first term is an attractive term representing the electrons on monomer A interacting with the nuclei on monomer B. The second term is a repulsive term where electrons on monomer B are interacting with electrons on monomer A via each corresponding charge distribution. The C_{AR} coefficients come from solving the coupled perturbed Hartree-Fock (CPHF) equations where $(A|\omega_B|R)$ is the perturbation. Because the induction is not solved self-consistently, any newly formed charge distributions that

can cause induction will be captured in the δHF term (that is solved self-consistently). The factor of two comes from the spin states for these one-electron excitations.

1.2.3.4 Dispersion:

$$Dispersion = 4 \sum_{ARBS} \Big[(AR|BS) / (\epsilon_A + \epsilon_B - \epsilon_R - \epsilon_S) * (AR|BS) \Big] \tag{42}$$

Dispersion (commonly referred to as the London dispersion force) is the physical quantity that typically gives computational chemists trouble because of the need for dynamic electron correlation to properly describe it. The minimum level required to adequately capture dispersion from first principles is second-order Møller-Plesset perturbation theory (MP2), and it can be noticed that equation 42 from SAPT looks very similar in spirit to the original MP2 equations. Equation 42 should be thought of as two separate pieces: $(AR|BS) / (\epsilon_A + \epsilon_B - \epsilon_R - \epsilon_S)$ represents the probability of the newly excited configuration being found; $(AR|BS)$ represents how stabilizing this new configuration is with respect to the total energy. The factor of four comes from this term being a two-electron term and the corresponding spin states.

1.2.4 Focal-Point Analysis

For a more detailed discussion of focal-point methods, please refer to the following references.[49, 37] Because of steep $\mathcal{O}(N^7)$ asymptotic scaling, where N represents the size of the system, CCSD(T) can only be applied to systems with a modest number of atoms (up to ∼30 atoms). To extend the range of applicability of large basis CCSD(T), a focal-point analysis[48, 38] approach is often used to estimate CCSD(T) in the large basis:

$$E^{large}_{CCSD(T)} \approx E^{large}_{MP2} + \delta^{CCSD(T)}_{MP2}, \tag{43}$$

$$\delta^{CCSD(T)}_{MP2} = E^{small}_{CCSD(T)} - E^{small}_{MP2}. \tag{44}$$

The subscripts and superscripts on energies refer to the method and basis set, respectively. The use of focal-point analysis has been applied to non-bonded complexes in many recent studies.[221, 199, 101, 82, 100, 103, 198, 145, 23, 95, 155, 134, 50] While such a focal-point approach has been shown to be reliable, caution still needs to be taken when selecting the "small" basis set for the $\delta_{MP2}^{CCSD(T)}$ correction, as shown in our recent work.[134]

1.3 Modern Approximations

1.3.1 Density Fitting

In electronic structure theory, the evaluation and storage of four-index integrals is a common bottleneck. Various approaches to this problem have been explored, such as resolution of the identity[235, 46, 47, 54, 21, 108, 226, 233, 230] [now commonly referred to as density fitting (DF)], Cholesky decompositions[19, 180, 117, 12, 24, 227, 10, 246, 25, 33] (CD), and pseudo-spectral[137, 136, 56] techniques. In the DF treatment, four-index integrals $(\mu\nu \mid \rho\sigma)$ are approximated by summations over three-index quantities:

$$(\mu\nu \mid \rho\sigma) \approx \sum_{PQ} (\mu\nu \mid P)[J^{-1}]_{PQ}(Q \mid \rho\sigma), \tag{45}$$

where $[J^{-1}]_{PQ}$ is the inverse of the Coulomb metric evaluated in an auxiliary basis set,

$$[J]_{PQ} = \int P(\mathbf{r}_1)\frac{1}{\mathbf{r}_{12}}Q(\mathbf{r}_2)d^3\mathbf{r}_1 d^3\mathbf{r}_2. \tag{46}$$

The three-index quantity $(\mu\nu \mid P)$ serves to cast the product $(\mu\nu \mid$ onto the auxiliary basis via the Coulomb metric

$$(\mu\nu \mid P) = \int \mu(\mathbf{r}_1)\nu(\mathbf{r}_1)\frac{1}{\mathbf{r}_{12}}P(\mathbf{r}_2)d^3\mathbf{r}_1 d^3\mathbf{r}_2. \tag{47}$$

While density-fitting does not lower the asymptotic scaling of MP2, it does reduce the prefactor significantly, with speed-ups in the range of 2 to 5.5 reported.[233, 193]

There exist many more methods for speeding up the evaluation of the correlation energy (e.g., local molecular orbital approaches such as local-MP2[233, 230]), yet the application of DF alone is often sufficient to reduce the cost of the correlation energy computation to the point that the time needed for the underlying self-consistent field (SCF) becomes the rate-determining step.

1.3.2 Dual-Basis

For a full review of dual-basis methods, please see references.[105, 238, 127, 206, 41] The dual-basis (DB) approximation proposed in the work of Steele *et al.* involves performing an iterative SCF in a small basis, followed by a single Roothan diagonalization step in a larger target basis set.[206] In practice, the small basis is typically a specially designed subset of the target basis set, although this restriction is not imposed by the theory. Once the SCF is converged with the small basis set, the occupied molecular orbital (MO) coefficients are projected onto the larger basis via

$$\mathbf{C}_{\overline{\mu}i} = \sum_{\overline{\mu\nu}} \sum_{\lambda} \mathbf{S}^{-1}_{\overline{\mu\nu}} \mathbf{S}_{\overline{\nu}\lambda} \mathbf{C}_{\lambda i}, \tag{48}$$

where $\mathbf{S}$ is the atomic orbital (AO) overlap matrix, i represents a MO index, Greek letters represent AO indices, and barred indices signify large-basis quantities. Using the newly constructed coefficient matrix, the new density matrix $\mathbf{P}$ is formed and a single Fock matrix is built and diagonalized. After including some first-order corrections, the DB-SCF energy is shown to be

$$E_{dual\ basis} = E_{small\ basis} + \sum_{\overline{\mu\nu}} \Delta\mathbf{P}_{\overline{\mu\nu}} \mathbf{F}_{\overline{\mu\nu}}, \tag{49}$$

where $\Delta\mathbf{P} = \mathbf{P}' - \mathbf{P}$ is the difference in the post-diagonalization density matrix $\mathbf{P}'$ and the small basis density matrix $\mathbf{P}$ (projected into the large basis). The small truncated basis sets used in the dual-basis methods have already been implemented[191] in the Q-Chem 3.2 program suite for several Pople and Dunning basis sets.

1.3.3 Explicitly Correlated Wavefunctions

For a detailed explanation of explicitly correlated wavefunction based methods, the reader is referred to the following references:[94, 122, 216, 153, 152, 151, 115, 224, 215, 55, 225, 113, 195, 220, 78]. For wavefunction based methods, one of the largest limitations remaining in computational chemistry is the very slow convergence of the correlation energy with respect to the size of the basis set employed. From studying complete basis set (CBS) extrapolations, it is known that the correlation energy converges close to l_{max}^{-3}, where l_{max} is the maximum angular momentum contained in the basis sets employed. While CBS extrapolation techniques do help the convergence problem to some extent, they really only work with relatively large angular momentum basis sets and introduce one or two parameters, though they are physically justified. The cause of the slow convergence is related to the presence of the correlation cusp in the wavefunction which requires that as two electrons converge, the wavefunction must be linear in r_{12}. The concept of explicitly correlated theories is to add corrections such that these electron-electron interactions are accounted for explicitly instead of trying to approximate them by including extremely large angular momentum basis sets. This concept was first demonstrated by Hylleraas[94] in 1929 on helium. The difficulty of this approach is that as you correlate N electrons explicitly, N-electron integrals are required. Computing all the N-electron integrals would restrict the molecular system to single atoms unless approximations are employed. Explicitly correlated versions of MP2 introduce three- and four-center integrals. Szalewicz *et al.* has shown that these can be avoided by using a weak orthogonality functional.[210] The remaining three-electron integrals have been computed employing the resolution of identity approximation.[122] Recently, short-range correlation factors such as $\exp(-\beta r_{12})$ have been shown to be more reliable than the original correlation factors which were linear in r_{12} (herein referred to as F12 and R12, respectively). In both theories, typical doubly excited configurations ϕ_{ij}^{ab}, like those

found in MP2, need to be augmented with explicitly correlated configurations ϕ_{ij}^{kl}, which will describe the electron-electron interactions near the correlation cusp,

$$|\phi_{ij}^{ab}\rangle = \hat{E}_{ai}\hat{E}_{bj}|\Psi^{(0)}\rangle, \tag{50}$$

$$|\phi_{ij}^{kl}\rangle = |\Phi_{ij}^{\alpha\beta}\rangle \mathcal{F}_{\alpha\beta}^{kl}, \tag{51}$$

and

$$\mathcal{F}_{\alpha\beta}^{kl} = \langle kl|\hat{F}_{12}\hat{Q}_{12}|\alpha\beta\rangle. \tag{52}$$

Since there are a very large number of possible doubly-excited configurations that could be explicitly correlated, $\mathcal{F}_{\alpha\beta}^{kl}$ can be employed to project down into a much smaller and more manageable set $|\phi_{ij}^{kl}\rangle$. The correlation factor $\hat{F}$ is defined to be either r_{12} or $\exp(-\beta r_{12})$, for R12 and F12 methods, respectively. The $\hat{Q}_{12}$ projector ensures strong orthogonality between the newly introduced excitations and the reference. Using these newly constructed doubly-excited configurations and handling all the electron-electron interactions explicitly in these new configurations leads to a correction to the second order energy as,

$$E^{(2)} = E_{MP2}^{(2)} + \Delta E_{F12}^{(2)}. \tag{53}$$

1.4 Organization of Thesis

This thesis contains six chapters starting with this introduction to electronic structure theory and more modern approximations that are employed by computational chemists. Chapter 2 will demonstrate how simple benchmarks can be used to understand new chemical phenomena. Chapter 3 will demonstrate how benchmark databases are developed as well as the errors we make in benchmark computations. Since our best benchmark quality computations are focal-point methods, both the MP2 component and the $\delta_{MP2}^{CCSD(T)}$ correction needed to be fully investigated. Chapter 4 provides an error and computational performance analysis of modern approximations applied to MP2. Chapter 5 (as well as Chapter 3) provide a very detailed

examination of the $\delta_{MP2}^{CCSD(T)}$ correction. Chapter 6 illustrates how to use benchmark quality data to develop new methods that are computationally less expensive without incurring large errors through empirical parameters. These chapters are adapted from previously published or submitted first author papers listed below:

Chapter 2: Marshall, M. S.,Steele, R. P.,Thanthiriwatte, K. S.,Sherrill, C. D., "Potential Energy Curves for Cation–π Interactions: Off-Axis Configurations Are Also Attractive," *J. Phys. Chem. A*, vol 113, p. 13628, 2009.

Chapter 3: Marshall, M. S., Burns, L. A., and Sherrill, C. D., "Basis set convergence of the coupled-cluster correction, $\delta_{MP2}^{CCSD(T)}$: Best practices for benchmarking non-covalent interactions and the attendant revision of the S22, NBC10, HBC6, and HSG databases," *J. Chem. Phys.*, vol. 135, p. 194102, 2011.

Chapter 4: Marshall, M. S.,Sears, J. S., Burns, L. A.,Brédas, J. L.,Sherrill, C. D., "An Error and Efficiency Analysis of Approximations to Møller–Plesset Perturbation Theory," *J. Chem. Theory Comput.*, vol. 6, p. 3681, 2010.

Chapter 5: Marshall, M. S. and Sherrill, C. D., "Averaging counterpoise- and uncounterpoise-corrected interaction energies and $\delta_{MP2}^{CCSD(T)}$ corrections for non-bonded complexes," *In preparation*

Chapter 6: Marshall, M. S. and Sherrill C. D., "Dispersion-Weighted Explicitly Correlated Coupled-Cluster Theory [DW-CCSD(T**)-F12]," *J. Chem. Theory Comput.*, vol. 7, p. 3978,2011

CHAPTER II

POTENTIAL ENERGY CURVES FOR CATION-π INTERACTIONS: OFF-AXIS CONFIGURATIONS ARE ALSO ATTRACTIVE

2.1 Abstract

Accurate potential energy surfaces for benzene·M complexes (M = Li^+, Na^+, K^+, and NH_4^+) are obtained using coupled-cluster theory through perturbative triple excitations, CCSD(T). Our computations show that off-axis cation-π interactions, where the cation is not directly above the aromatic ring, can be favorable and may influence molecular recognition. Even perpendicular, side-on interactions retain 18-32% of their π-face interaction energy in the gas phase, making their bond strengths comparable to hydrogen bonds in the gas phase. Solvent effects have been explored for each complex using the polarizable continuum model.

2.2 Introduction

Cation-π interactions are one of the fundamental forces of molecular recognition,[65] and they have been implicated in a number of biochemical processes, including the binding of acetylcholine by various acetylcholine receptors, ion selectivity in K^+ channels, and steroid bio-synthesis.[72, 44, 244] Meadows *et al.* estimate that 8% of protein residues are potentially involved in cation-π interactions.[140] The large number of cation-π interactions present in proteins argues for a more complete understanding of the nature of these interactions. Cation-π interactions have been the focus of a vast array of recent research efforts.[196, 242, 71, 106, 241, 119, 147, 202, 222, 34, 167, 173, 171, 36, 65, 148, 142, 20, 140, 6, 57, 52, 146, 129, 44, 141, 32, 27, 107]

The attraction between a cation and an aromatic ring can be quite strong. Although nonpolar, benzene can compete with a water molecule for binding a cation in the gas phase. For example, the interaction energy of benzene·K^+ is -18 kcal mol^{-1}, compared to -19 kcal mol^{-1} for $K^+(H_2O)$.[208] The experimental interaction energies of benzene·Na^+ and benzene·Li^+ are stronger still, at -22.13±1.39 and -38.50±3.23 kcal mol^{-1}, respectively.[6]

Understanding the nature of the cation-π interaction in model systems has been a topic of much recent interest. Earlier studies explained the strength of the cation-π interaction as arising primarily from the electrostatic attraction between an ion and the large quadrupole moment featured by many aromatic rings.[44] However, it is now clear that the induction term is at least as important as the electrostatic term.[202, 222, 223]

Very high-level quantum mechanical computations on the prototype benzene·Na^+ complex have been reported by Feller,[52] who estimated interaction energies using coupled-cluster theory with perturbative triple excitations [CCSD(T)] and complete basis set (CBS) extrapolations. Feller's theoretical interaction energy (ΔH_0 = -24.4 ± 0.3 kcal mol^{-1}) falls in between two experimental results (-21.1 and -27.6 kcal mol^{-1})[13] which differ by 6.5 kcal mol^{-1}. An additional high-level theoretical study by Feller, Dixon, and Nicholas[53] for the complexes of alkali earth cations with benzene indicates that lower-level computations can yield 298 K binding energies which are 3-4 kcal mol^{-1} too low. Woolf and co-workers have found [239] that the CHARMM force field does not always reliably model the interaction of indoles with point charges when compared to the more reliable density functional theory (DFT).

The effect of substituents on cation-π interactions has been examined in several recent theoretical and experimental papers.[141, 7, 8, 45] Electrostatic interactions appear to be sufficient to explain computed trends in substituent effects, [141, 7, 8] although Amunugama and Rodgers also point out the role of the polarizability of the

aromatic.[7, 8] The interaction of benzene with ions MX^+ has also been examined,[31] where M is an alkaline earth dication and X is a counter ion (e.g., H^-, CH_3^-).

Despite the large number of recent studies on cation-π interactions, previous work focuses only on cations interacting directly with the face of the aromatic ring. This is rather surprising considering that many off-angle cation-π interactions have been found in proteins.[36, 57, 146, 32] Because the geometry of cation-π interactions may be constrained when these interactions occur in larger chemical or biochemical systems, it is important to know how the attractions behave as a function of distance and orientation. While these interactions will certainly be attenuated in larger systems due to polarization of the environment, the gas phase potential energy curves represent the first step in understanding the nature of the fundamental cation-π interaction. Moreover, they may serve as benchmarks in the development of the next generation of polarizable force fields which may accurately model noncovalent interactions in biochemical systems. In this work, we present high-level quantum mechanical computations of potential energy curves for cation-π interactions by considering complexes of the cations Li^+, K^+, Na^+, and NH_4^+ with benzene, and we include configurations in which the cation is not directly above the aromatic ring. Contrary to expectations based on electrostatic interactions alone, we find that off-axis approaches can be favorable—although not nearly as favorable as binding to the top of the ring—suggesting that off-axis cation-π interactions should also be considered when analyzing biochemical systems.

2.3 Theoretical Approach

Potential energy surfaces (PES) were computed using the CCSD(T) method[172] with the Pople 6-311++G(2d,2p) basis set.[120, 139, 22] All interaction energies were corrected for basis set superposition error (BSSE) using the scheme outlined by Boys and Bernardi.[26] All single point energies had the following orbitals frozen for each

atom: C[1s], Na[1s], K[1s,2s,2p], N[1s], Li[none]. The curves were computed with rigid monomers employing the benzene geometry recommended by Gauss and Stanton: $r_{CC} = 1.3915$Å and $r_{CH} = 1.0800$Å.[58] The NH_4^+ geometry, $r_{NH} = 1.0235$Å and $\theta_{HNH} = 109.467^o$, was acquired from a full CCSD(T)/aug-cc-pVTZ optimization. The NH_4^+ is oriented such that only one hydrogen points to the center of the ring. Very small changes in the interaction energy are found for different orientations. Displacements for the PES are calculated from the center of mass of the benzene to the center of mass of the cation. Rigid PES scans were performed on a dense grid of points for R $\in [2.0\text{Å}, 7.0\text{Å}]$, $\theta \in [0°, 90°]$, and $\phi = 0°$ or $30°$ (see Figure 3).

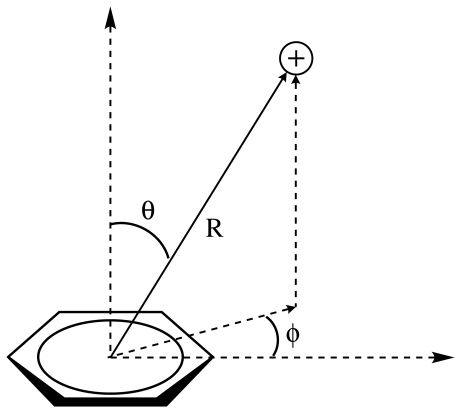

Figure 3: Geometries for cation-π systems.

All computations were carried out with the MOLPRO 2006 program[232]. Because we are tracing potential curves and because monomer geometries were frozen, the geometries considered do not represent stationary points on the potential surface. Hence, zero point corrections were not applied.

2.4 Results and Discussion

Cation-π Interactions

Potential energy surfaces for various metal cations and ammonium interacting with aromatic benzene are shown in Figure 4.

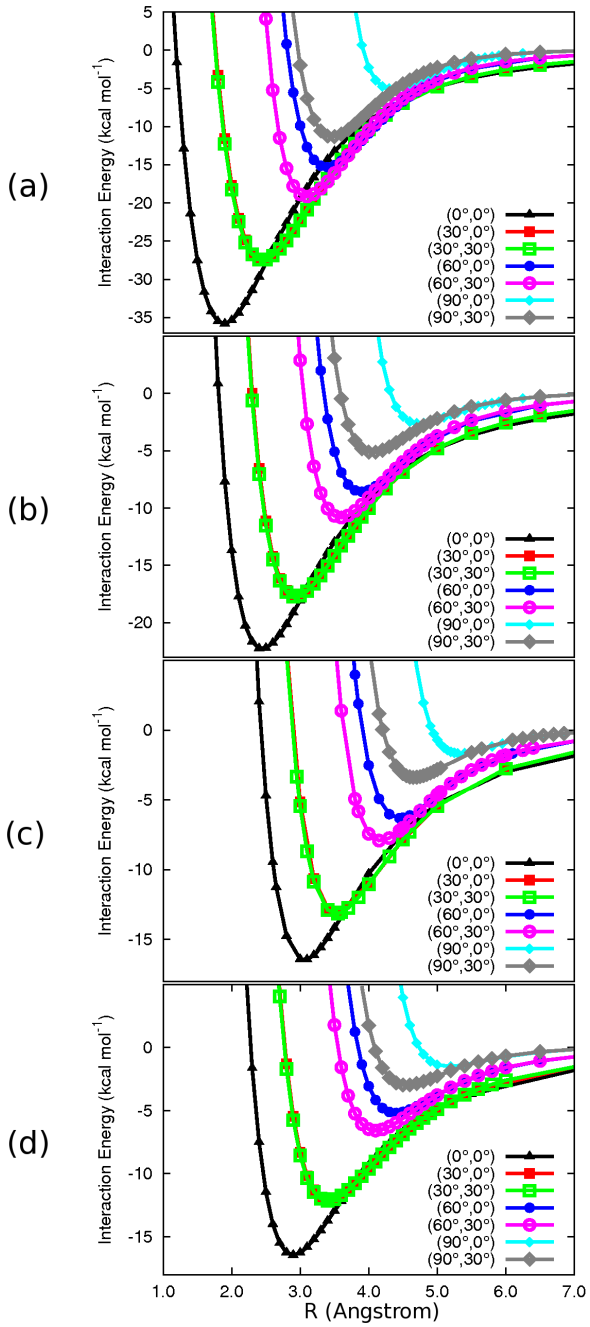

Figure 4: Potential energy curves for given (θ,ϕ) values for benzene·M^+ (M = Li, Na, K, NH_4) at the CCSD(T)/6-311++G(2d,2p) level of theory. (a) benzene·Li^+ (b) benzene·Na^+ (c) benzene·K^+ (d) benzene·NH_4^+

Interaction energies for the minimum of each curve are given in Table 1.

Table 1: CCSD(T)/6-311++G(2d,2p) Counterpoise Corrected Interaction Energies (E_{int}, kcal mol^{-1}) and Equilibrium Distances (R, Angstrom)

Cation	θ	ϕ	E_{int}	R
Li^+	0	0	-35.8	1.9
	30	0	-27.3	2.5
	60	0	-15.3	3.4
	90	0	-5.0	4.3
	30	30	-27.4	2.4
	60	30	-19.2	3.1
	90	30	-11.3	3.5
Na^+	0	0	-22.2	2.4
	30	0	-17.6	2.9
	60	0	-8.6	3.9
	90	0	-2.6	4.7
	30	30	-17.7	2.9
	60	30	-10.8	3.6
	90	30	-5.1	4.1
K^+	0	0	-16.5	2.9
	30	0	-12.1	3.4
	60	0	-5.2	4.4
	90	0	-1.5	5.2
	30	30	-12.1	3.4
	60	30	-6.6	4.1
	90	30	-3.0	4.5
NH_4^+	0	0	-16.4	3.1
	30	0	-13.0	3.6
	60	0	-6.3	4.5
	90	0	-1.7	5.3
	30	30	-13.1	3.6
	60	30	-7.9	4.2
	90	30	-3.4	4.6

In each case, the most favorable geometry is one where the cation is directly above the ring (θ=ϕ=0°). For this configuration, the equilibrium CCSD(T)/6-311++G(2d,2p) interaction energies are -35.8, -22.2, -16.5, and -16.4 kcal mol^{-1} for Li^+, Na^+, K^+, and NH_4^+ respectively. Our results for the cation above the ring ($\theta = 0°$) are in good

agreement with the previous experimental work of Amicangelo *et al*[6] who report interaction energies of -38.5, -22.1, and -17.5 kcal mol^{-1} for Li^+, Na^+, and K^+ respectively, even though our results are purely electronic binding energies, neglecting enthalpy corrections. Our results are slightly less bound than the very high-quality CCSD(T) complete-basis-set estimates of Feller *et al.*,[53] who obtain electronic interaction energies of -38.0, -25.4, and -20.6 kcal mol^{-1} for Li^+, Na^+, and K^+. We attribute the differences primarily to remaining deficiencies in the basis set and to the lack of geometry relaxation in our results (full geometry relaxation is somewhat more important for these cation-π interactions than for neutral non-covalent interactions, and it can be as large as $\sim$ 1.5 kcal mol^{-1} for benzene·Li^+).

As the cation is moved from above the aromatic face ($\theta = 0°$) to the side-on geometry ($\theta = 90°$), we see that up to 30% of its interaction energy is retained, meaning the side-on interactions are still significantly stabilized (-11.33, -5.13, -2.96, and -3.38 kcal mol^{-1} for Li^+, Na^+, K^+, and NH_4^+ respectively, at the optimum values of R). The side-on geometries are saddle points, rather than local minima, on the gas phase potential energy surface. Nevertheless, side-on or near side-on geometries can occur in complex environments such as proteins due to backbone or other steric constraints. Although the gas phase interaction energies would be reduced in solution, their large magnitude indicates that even side-on interactions may be important in molecular recognition.

For all configurations where the cation is not directly above the benzene, there was a clear preference for $\phi = 30°$ (where the cation is positioned between two H atoms) over $\phi = 0°$ (where the cation is aligned in a vertical plane with an H atom; see Figure 3). Our results indicate that the size of the cation plays a substantial role in its interaction energy. As size of the cation increases, the interaction energy decreases. This is primarily due to the significant role of induction. The small cations can approach closer to the ring and therefore can induce a much larger polarization

response from the benzene.

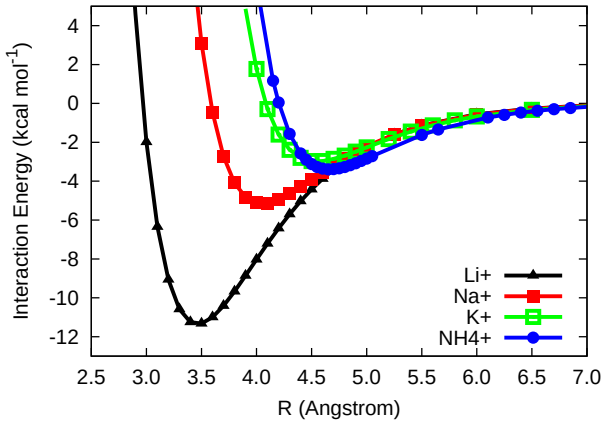

Figure 5: Potential energy curves for all cations at θ=90°,ϕ=30°

Figure 5 compares the potential energy curves for the side-on configurations (θ = 90°, ϕ = 30°) for the various cations considered. The curves for Na^+, K^+, and NH_4^+ are similar, but Li^+ is significantly more bound because of its smaller size.

Decomposition of Na^+-π interaction energy

To better understand how the cation-π interaction changes from in-plane (θ = 90°,ϕ = 30°) to above the π-face (θ = 0°, ϕ = 0°) we utilize density functional based symmetry adapted perturbation theory (DFT-SAPT) [96] to examine benzene·Na^+ as a representative example. DFT-SAPT is a recent extension of SAPT[98, 237] theory where the interacting monomers are expressed in terms of Kohn-Sham DFT. DFT-SAPT has been shown to accurately reproduce coupled-cluster interaction energies.[75] In order to achieve such accuracy, one must asymptotically correct the exchange-correlation functional.[73, 74, 76] In this work we employ the gradient-regulated asymptotic correction of Gruning *et al.*[66] The shift parameters used for benzene and Na^+ are 0.0713 and 0.2783 hartree respectively. These values are calculated from the difference between the HOMO energy and the exact ionization potential of each monomer. The exact ionization potentials are taken from Cohen *et al.*[35] for Na^+

(47.2864 eV) and Lias *et al.*[128] for benzene (9.2459 eV). All DFT-SAPT calculations were carried out with the MOLPRO 2006 program[232] using the PBE0 density functional [3] and the 6-311++G(2d,2p)[120, 139, 22] basis set.

Within the DFT-SAPT framework, the interaction energy is decomposed into the following contributions: electrostatics, induction, exchange, and dispersion. The total interaction energy can be defined as follows:

$$E_{interaction}^{DFT-SAPT} = E_{elst}^{(1)} + E_{exch}^{(1)} + E_{ind}^{(2)} + E_{exch-ind}^{(2)} + E_{disp}^{(2)} + E_{exch-disp}^{(2)} \tag{54}$$

The superscripts refer to the order of the correction. As done in our previous studies and by others,[196, 86] we collect energy components as follows:

$$E_{ind} = E_{ind}^{(2)} + E_{exch-ind}^{(2)} \tag{55}$$

$$E_{disp} = E_{disp}^{(2)} + E_{exch-disp}^{(2)} \tag{56}$$

which results in an total interaction energy of

$$E_{total} = E_{exch}^{(1)} + E_{elst}^{(1)} + E_{disp} + E_{ind} \tag{57}$$

Using the methods and decomposition scheme outlined above, the energetic contributions for the above-face and in-plane configurations are computed for benzene·Na^+ (Figure 6). For the cation above the π-face, the interaction is equally stabilized by electrostatics (47%) and induction (49%), with dispersion only contributing 4% of the stabilizing interaction. Both CCSD(T) and DFT-SAPT predict that the in-plane configuration are bound. With the exception of the small contribution of dispersion, the in-plane configuration is entirely bound by induction. The electrostatic term for the in-plane configuration is repulsive because the closest contacts are between the cation and two hydrogens of benzene, which have partial positive charges. Most likely this is the reason earlier studies disregarded these configurations. To validate the reliability of DFT-SAPT for cation-π interactions, we compare DFT-SAPT to the

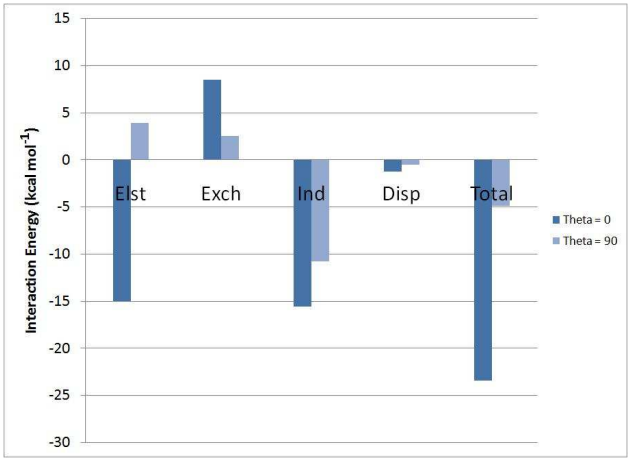

Figure 6: DFT-SAPT decomposition of benzene·Na^+

published SAPT data of Soteras *et al.*[202] for the case of Na^+ above benzene (see Figure 7). Note also how well the interaction energy predicted by DFT-SAPT (-4.9 kcal mol^{-1}) compares to CCSD(T) (-5.1 kcal mol^{-1}).

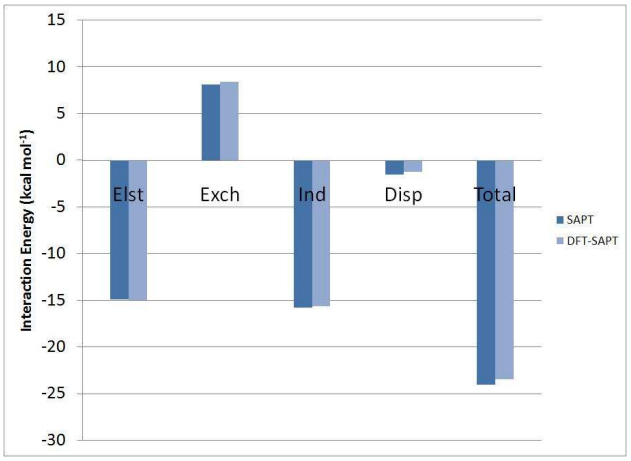

Figure 7: DFT-SAPT decomposition of benzene·Na^+ compared to wavefunction based SAPT for θ=0° (cation above the ring)

Solvent Effects

Solvent effects on the strength of the cation-π interactions were examined with polarized continuum model (PCM)[144, 219, 30, 181] using the B3LYP[18] density

functional and the 6-311++G(2d,2p) basis set. All solvent calculations were performed using the GAMESS program.[184] The CCSD(T) interaction energies were corrected by adding the solvent free energy of interaction (ΔG^{int}_{solv}) using equations 58 and 59.

$$E^{solv-int}_{total} = E^{int}_{CCSD(T)} + \Delta G^{int}_{solv} \tag{58}$$

where

$$\Delta G^{int}_{solv} = \Delta G^{AB}_{solv} - \Delta G^{A}_{solv} - \Delta G^{B}_{solv} \tag{59}$$

Three configurations were considered for each cation-benzene complex [θ,ϕ=(0,0), (60,30), and (90,30)] and R taken from equilibrium distance computed using CCSD(T)/6-311++G(2d,2p) level of theory. Interaction energies were computed for water (ϵ=78.39) and chloroform (ϵ=4.9). Chloroform was chosen because it has a dielectric constant closer to what one would find in the interior of a protein.[149] Solvated interaction energies are shown in Table 2.

Table 2: Solvated Interaction Energies in kcal mol^{-1}. Gas phase computed at CCSD(T)/6-311++G(2d,2p). Solvation correction computed at B3LYP-PCM/6-311++G(2d,2p).

Cation	θ	ϕ	R (Ang)	Gas phase	Chloroform	Water
Li^+	0	0	1.9	-35.8	-21.2	-18.4
	60	30	3.1	-19.2	-8.0	-6.0
	90	30	3.5	-11.3	0.7	-0.7
Na^+	0	0	2.4	-22.2	-15.1	-14.4
	60	30	3.6	-10.8	-4.8	-4.2
	90	30	4.1	-5.10	0.4	0.9
K^+	0	0	2.9	-16.5	-9.2	-8.5
	60	30	4.1	-6.6	-1.2	-0.8
	90	30	4.5	-3.0	1.5	1.6
NH_4^+	0	0	3.1	-16.4	-2.6	-0.4
	60	30	4.2	-7.9	2.3	3.6
	90	30	4.6	-3.4	3.9	4.6

attenuated by solvent, they can still be favorable, even for some of the off-axis configurations.

Comparison to lower levels of theory

Considering that lower levels of theory have been popular for studies of cation-π interactions,[167, 107, 147] it is useful to evaluate the reliability of more approximate methods for these systems. Møller-Plesset perturbation theory (MP2) (Figure 8) and Hartree-Fock self-consistent-field (SCF) (Figure 9) are compared to CCSD(T).

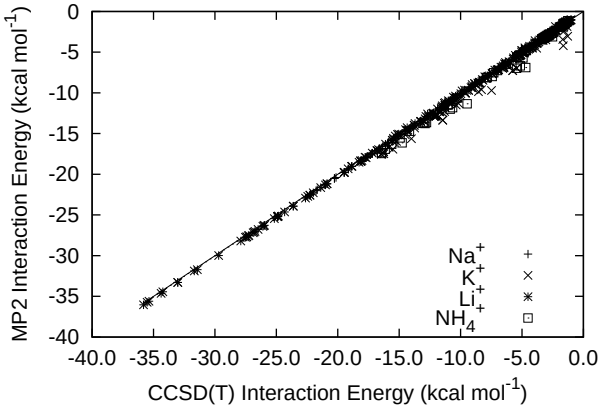

Figure 8: Comparison of MP2 to CCSD(T) Counterpoise Corrected Interaction Energies.

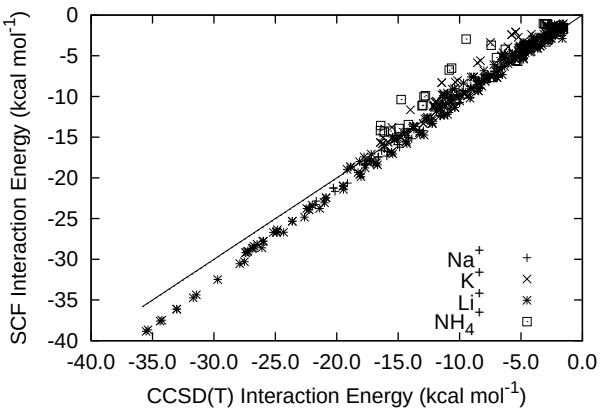

Figure 9: Comparison of SCF to CCSD(T) Counterpoise Corrected Interaction Energies.

Out of the geometries considered above, this analysis included all points where both levels of theory predicted binding. MP2 interaction energies are comparable to CCSD(T) with a root mean square deviation (RMSD) of only 0.43 kcal mol^{-1} (max deviation of 2.58 kcal mol^{-1}). SCF, on the other hand, has a RMSD of 1.25 kcal mol^{-1} (max deviation of 6.52 kcal mol^{-1}). SCF does not include dynamical electron correlation, therefore it fails to describe dispersion interactions. While not as dominant as the electrostatic and induction contributions, the dispersion term is expected to account for 2%, 6%, and 16% of the attractive interactions for Li^+, Na^+, K^+-benzene complexes (where the cation is directly above the π face), respectively.[202] If MP2 or CCSD(T) are not computationally affordable, the inclusion of an empirical dispersion term should be considered for methods like SCF or DFT. Density fitted MP2 (DF-MP2) has also been shown to be a very good substitute for MP2 for cation-π interactions.[171]

2.5 Conclusions

To our knowledge, this work is the first to present potential energy surfaces for cation-π interactions where the cation is not necessarily above the center of the aromatic ring.

Our computations at the CCSD(T)/6-311++G(2d,2p) level of theory show that even in-plane cation-π interactions can be favorable, with a gas-phase interaction energies of -3 to -11 kcal mol^{-1} for the systems considered here. These off-axis interactions are not as strong as those directly above the ring face, but are still near the strength of hydrogen bonds in the gas phase. Thus, we believe that the concept of a cation-π interaction should be broadened to include configurations where the cation may not be directly above the aromatic ring (and may even be in the same plane). PCM computations of solvent effects indicate that solvents reduce the strength of the cation-π interactions, but the off-axis configurations can still be attractive.

Because of the large induction contributions present in cation-π systems, many pairwise potentials currently used for large-scale simulations are not able to properly describe this type of bonding. It is our hope that the potential curves presented here may be used as valuable data for the calibration of next generation polarizable force fields.

CHAPTER III

BASIS SET CONVERGENCE OF THE COUPLED-CLUSTER CORRECTION, $\delta_{MP2}^{CCSD(T)}$: BEST PRACTICES FOR BENCHMARKING NON-COVALENT INTERACTIONS AND THE ATTENDANT REVISION OF THE S22, NBC10, HBC6, AND HSG DATABASES

3.1 Abstract

In benchmark-quality studies of non-covalent interactions, it is common to estimate interaction energies at the complete basis set (CBS) coupled-cluster through perturbative triples [CCSD(T)] level of theory by adding to CBS second-order perturbation theory (MP2) a "coupled-cluster correction," $\delta_{MP2}^{CCSD(T)}$, evaluated in a modest basis set. This work illustrates that commonly used basis sets such as 6-31G*(0.25) can yield large, even wrongly signed, errors for $\delta_{MP2}^{CCSD(T)}$ that vary significantly by binding motif. Double-ζ basis sets show more reliable results when used with explicitly correlated methods to form a $\delta_{MP2-F12}^{CCSD(T^*)-F12}$ correction, yielding a mean absolute deviation of 0.11 kcal mol^{-1} for the S22 test set. Examining the coupled-cluster correction for basis sets up to sextuple-ζ in quality reveals that $\delta_{MP2}^{CCSD(T)}$ converges monotonically only beyond a turning-point at triple-ζ or quadruple-ζ quality. In consequence, CBS extrapolation of $\delta_{MP2}^{CCSD(T)}$ corrections before the turning point, generally CBS(aug-cc-pVDZ,aug-cc-pVTZ), are found to be unreliable and often inferior to aug-cc-pVTZ alone, especially for hydrogen-bonding systems. Using the findings of this paper, we revise some recent benchmarks for non-covalent interactions, namely the S22, NBC10, HBC6, and HSG test sets. The maximum differences in the revised benchmarks are

0.080, 0.060, 0.257, 0.102 kcal mol^{-1}, respectively.

3.2 Introduction

High-accuracy benchmark data sets have become a cornerstone for testing new theories, basis sets, and approximations in computational chemistry. As the field becomes more reliant on these high-quality test sets, a better understanding of their underlying errors is required. One common approach for obtaining benchmark quality reference data is through focal-point analysis.[48, 38] In the context of non-covalent interactions, focal-point analysis is often used to estimate coupled-cluster theory through perturbative triple excitations in a large basis set [est. CCSD(T)/large]:

$$E^{large}_{CCSD(T)} \approx E^{large}_{MP2} + \delta^{CCSD(T)}_{MP2}, \tag{60}$$

$$\delta^{CCSD(T)}_{MP2} = E^{small}_{CCSD(T)} - E^{small}_{MP2}. \tag{61}$$

The subscripts and superscripts on energies refer to the method and basis set respectively. The coupled-cluster correction $\delta^{CCSD(T)}_{MP2}$ is also referred to as ΔCCSD(T) in the literature. The "large" basis set is often a complete basis set (CBS) extrapolation, whereas the "small" basis set used for $\delta^{CCSD(T)}_{MP2}$ is usually a single basis set, though it may also be a basis set extrapolation.[193] As has been noted,[199, 101, 82, 198, 155] this focal-point approach works well because even though the convergence rates of CCSD(T) and MP2 correlation energies are slow with respect to basis set size, the rate of convergence of the difference [CCSD(T)–MP2] is much faster.[100] Thus, a much smaller basis set may be used for the $\delta^{CCSD(T)}_{MP2}$ correction than for the underlying SCF and MP2 computations. This approach has been applied to obtain several non-covalent interaction benchmark energies.[221, 199, 103, 50] One of the more common small basis sets used is 6-31G*(0.25), particularly for benchmarking

non-covalent interactions in bio-molecules. This modified Pople basis set is formed by replacing the usual exponent for the d polarization functions with a more diffuse exponent (α_d=0.25) better able to describe nonbonding interactions.[121, 81, 84, 83] While this approach has been used widely, there now exist several papers showing that $\delta_{MP2}^{CCSD(T)}$ corrections with modest basis sets (e.g., 6-31G*, cc-pVDZ, aug-cc-pVDZ) often lead to inaccurate interaction energies. One such paper by Boese *et al.*[23] reports interaction energy errors of 10% for neutral hydrogen-bonded complexes using 6-31G*(0.25) and claims that using such small basis sets "does more harm than good" (because too small a basis set can yield the wrong sign for the $\delta_{MP2}^{CCSD(T)}$ correction). A large error is also seen in the work of Min *et al.* on benzene·Na^+.[145] They report a $\delta_{MP2}^{CCSD(T)}$ correction increasing in magnitude from -0.13 kcal mol^{-1} in the aug-cc-pVDZ basis to -1.16 kcal mol^{-1} in the aug-cc-pVTZ basis. Pitoňák *et al.*[155] recently studied the effect of small basis sets on stacked adenine·thymine. They find that small Pople basis sets have errors less than 10-20% in the $\delta_{MP2}^{CCSD(T)}$ correction. While this nucleobase test system is interesting, it is not representative of difficult cases in which the $\delta_{MP2}^{CCSD(T)}$ correction is a significant portion of the binding energy. In stacked adenine·thymine, $\delta_{MP2}^{CCSD(T)}$ is 3.18 kcal mol^{-1} compared to a total interaction energy of -11.66 kcal mol^{-1}; whereas, the more difficult cases are systems like the parallel-displaced benzene dimer where the $\delta_{MP2}^{CCSD(T)}$ correction is 1.96 kcal mol^{-1} compared to a total interaction energy of only -2.67 kcal mol^{-1}.[193] When the $\delta_{MP2}^{CCSD(T)}$ correction is large compared to the interaction energy, a significant error in this quantity can result in a large relative error in the total binding of the complex. This is yet another reason why the benzene dimer has been examined in such detail.[199, 221, 197, 160, 95, 157, 60, 193] In particular, a key study by Janowski and Pulay[95] demonstrated that even the reasonably good aug-cc-pVDZ basis set is not quite sufficient to converge $\delta_{MP2}^{CCSD(T)}$ within 0.1 kcal mol^{-1} of its true value for the benzene dimer. Given that some approximate methods for non-covalent interactions

are now achieving mean absolute deviations (MADs) of only few tenths of one kcal mol^{-1},[61, 29] it is important to begin considering how the remaining errors in the benchmark interaction energies can be reduced to only a few hundredths of one kcal mol^{-1}. In many cases, the $\delta_{MP2}^{CCSD(T)}$ term appears to be the largest remaining source of error, and hence it is the focus of the present study. In particular, we examine the basis set convergence of this correction for several small van der Waals dimers, and from this work we present revised benchmark interaction energies for several databases of non-covalent interactions.

While estimates of CCSD(T)/CBS interaction energies using a relatively small basis set to evaluate $\delta_{MP2}^{CCSD(T)}$ can be quite accurate in many cases, the quality of the procedure is not necessarily consistent across binding motifs for non-covalent interactions. This study examines the error incurred by using double-ζ basis set $\delta_{MP2}^{CCSD(T)}$ corrections for complexes that are hydrogen-bonded, dispersion-bound, or of mixed character. Particularly, we focus on the S22 benchmark set,[103] which features diverse types of non-bonded interactions over a wide range of system sizes, from water dimer (6 atoms) to adenine·thymine complexes (30 atoms).

This work also investigates the use of explicitly correlated wavefunctions to obtain better $\delta_{MP2}^{CCSD(T)}$ corrections at reduced computational cost. Analogously to Equations 70 and 71:

$$E_{CCSD(T^*)-F12}^{large} \approx E_{MP2-F12}^{large} + \delta_{MP2-F12}^{CCSD(T^*)-F12}, \tag{62}$$

$$\delta_{MP2-F12}^{CCSD(T^*)-F12} = E_{CCSD(T^*)-F12}^{small} - E_{MP2-F12}^{small}. \tag{63}$$

Explicitly correlated wavefunctions have been shown to produce accurate energies using relatively small basis sets.[230, 131, 231, 229, 4, 116] This work evaluates whether this same approach can more quickly converge the coupled-cluster correction.

3.3 Theoretical Approach

3.3.1 Notation

The Dunning basis sets aug-cc-pVNZ (N = D,T,Q,5,6) are herein referred to as aNZ. The heavy-aug-cc-pVNZ (N = D,T,Q,5,6), which is aug-cc-pVNZ on the non-hydrogen atoms and cc-pVNZ on the hydrogens, is herein referred to as haNZ. It is also worth noting here that none of the interaction energies presented in this work are at fully optimized minima, but instead at fixed, near-equilibrium geometries defined by the corresponding original papers. All interaction energies are counterpoise-corrected for basis-set superposition error (BSSE) with the scheme outlined by Boys and Bernardi.[26]

3.3.2 Convergence of CBS Extrapolated $\delta_{MP2}^{CCSD(T)}$ corrections

A better understanding is required of the convergence trends of $\delta_{MP2}^{CCSD(T)}$ corrections with respect to basis size for both individual basis sets (e.g. aNZ, N = D,T,Q,5,6) and 2-point Helgaker extrapolations[67] [e.g., CBS(aNZ,aMZ), NM = DT, TQ, Q5, 56]. To accomplish this, $\delta_{MP2}^{CCSD(T)}$ quantities are computed for each of the aforementioned basis sets for several of the smallest members of the S22 test set.

3.3.3 Revision of Benchmark Databases

In accordance with the conclusions of this study (discussed below), the reference interaction energies for the S22,[103] NBC10,[193, 213, 87] HBC6,[218] and HSG[50] databases have been revised, with geometries remaining unchanged. Benchmark values for the latter two have been computed as a sum of the HF/aQZ energy, the two-point (aTZ and aQZ) Helgaker CBS extrapolation[67] of the MP2 correlation energy, and the $\delta_{MP2}^{CCSD(T)}$ correction with the aTZ basis set (haTZ for HSG). Similarly, for the NBC10 test set, previous benchmark energies employing (h)aDZ/(h)aTZ extrapolations of $\delta_{MP2}^{CCSD(T)}$ have been updated to use simple (h)aTZ values of $\delta_{MP2}^{CCSD(T)}$. The databases at this revision level will be denoted as NBC10A, HBC6A, and HSG-A.

For the S22 test set, revisions of the original[103] benchmark energies have already been published. A recent paper by Takatani *et al.*[211] contributed high-quality estimates of the $\delta_{MP2}^{CCSD(T)}$ correction to define the S22A binding energies. The $\delta_{MP2}^{CCSD(T)}$ corrections were extrapolated to the CBS limit with an aDZ,aTZ 2-point Helgaker extrapolation[67] of the MP2 and CCSD(T) correlation energies. Another work independently revised binding energies for the S22 dimers using larger basis MP2/CBS energies and single-basis $\delta_{MP2}^{CCSD(T)}$ corrections with mid-bond functions.[161] These two studies agree within 0.044 kcal mol^{-1} averaged across the entire set, or within 0.029 kcal mol^{-1} if the adenine·thymine complexes (numbered 7 and 15) are dropped. Our group recently has analyzed the differences in these two benchmarks and concluded that aDZ,aTZ CBS extrapolated $\delta_{MP2}^{CCSD(T)}$ corrections typically slightly overestimate $\delta_{MP2}^{CCSD(T)}$ corrections for hydrogen bonded complexes. A detailed study of this is presented in Section III B. Additionally, we have performed new CCSD(T) computations using larger basis sets when feasible. The best quality results from among the literature values and new computations have been judiciously selected to form the S22B benchmark set. The S22 set conveniently partitions complexes by binding type into hydrogen bonding, dispersion-dominated, and mixed influence categories. This grouping allows one to see if particular methods struggle for certain kinds of non-covalent interactions. Interaction energy decompositions via DFT-SAPT[61] by Grafová *et al.* and SAPT2+(3)/aTZ results by Hohenstein and Sherrill[85] have shown that the original, intuitive assignments of the S22 complexes to binding-type subgroups were not wholly consistent. In this work, we use the grouping suggested by SAPT2+(3) data which moves stacked adenine·thymine (15) and uracil (13) to the mixed influence subset and T-shaped benzene dimer (20) to the dispersion-dominated subset. We note that the "hydrogen bonding" group might more precisely be designated as "electrostatically dominated," as the electrostatic character is what we confirmed by SAPT analysis. Some members of this group, such as NH_3 dimer, may

not necessarily fit the latest IUPAC definition of hydrogen-bonding.[150, 14]

All subsequent discussions employ the revised S22B, NBC10A, HBC6A, and HSG-A interaction energies as benchmarks.

3.3.4 Small Basis Set $\delta_{MP2}^{CCSD(T)}$ Corrections

For each of the complexes in the S22 test set, we report $\delta_{MP2}^{CCSD(T)}$ corrections (using equation 71) for the following basis small sets: 6-31G*, 6-31G*(0.25), 6-31G**(0.25,0.15), cc-pVDZ, and aug-cc-pVDZ (aDZ). The number in parentheses indicates the non-standard exponent on the polarization functions. These modified Pople basis sets were chosen because they are commonly used in the literature for computing $\delta_{MP2}^{CCSD(T)}$.[156, 104, 159, 39, 243, 174, 183, 182] Results are compared to the best currently available values of $\delta_{MP2}^{CCSD(T)}$ (several of which are revised in this work). Mean absolute deviation as well as mean absolute percent deviation (MAPD, with respect to the total interaction energy) are reported for each basis set. Considering the minimal role of core correlation towards overall interaction energies,[23] the frozen-core approximation was employed for all computations. All total energies were converged to 10^{-9} hartree.

3.3.5 Explicitly Correlated $\delta_{MP2-F12}^{CCSD(T^*)-F12}$ corrections.

Using equations 62 and 63, we examine the performance of explicitly correlated F12 methods for computing $\delta_{MP2}^{CCSD(T)}$ corrections. Some technical aspects of CCSD(T)-F12 require elaboration, the first of which is how to handle the perturbative triples correction. This work follows the approach of Marchetti *et al.*[132] whereby the triples correction is scaled by the ratio of MP2 correlation energy and MP2-F12 correlation energy:

$$E_{(T^*)}^{corr} = E_{(T)}^{corr} \cdot \frac{E_{MP2-F12}}{E_{MP2}}. \tag{64}$$

This procedure leads to a better triples correction, but if it is done independently for the dimer and each monomer, size-consistency is lost, as pointed out by Marchetti *et al.*[132] To retain size-consistency while computing an interaction energy, the scaling factor must be kept consistent for each complex. Any of the three scaling factors (dimer, monomer A, monomer B) could be chosen, though the dimer is the most common choice. Methods that employ a single scaling factor for each of the three computations are herein referred to as CCSD(T**)-F12. CCSD(T*)-F12 (with one asterisk) here designates independently scaled triples corrections. The second issue that must be considered is the choice of F12 ansatz. For CCSD-F12, we present both F12a and F12b.[4, 116] For MP2-F12, we present only MP2-F12/3C(FIX) as the MP2 reference.[229]

3.4 Results and Discussion

3.4.1 CBS Extrapolated $\delta_{MP2}^{CCSD(T)}$ Corrections

First, it is worthwhile to establish the best possible benchmark values for $\delta_{MP2}^{CCSD(T)}$. In previous work we presented a basis-set-consistent revision of interaction energies for the S22 test set, which we designated S22A.[211] Based on limited comparisons where extrapolated CCSD(T)/CBS(aTZ,aQZ) interaction energies were then available, it appeared that using CBS(aDZ,aTZ) $\delta_{MP2}^{CCSD(T)}$ corrections was very similar to and in some cases slightly preferable to using aTZ $\delta_{MP2}^{CCSD(T)}$ corrections. However, in the present work, we find that this previous picture is somewhat misleading—surprisingly—because the CCSD(T)/CBS(aTZ,aQZ) interaction energies are themselves not always fully converged.

Figure 10 illustrates the slow, non-monotonic convergence characteristic of the $\delta_{MP2}^{CCSD(T)}$ term, which often results in significant errors for estimates involving double-ζ values [i.e., aDZ and CBS(aDZ,aTZ)].

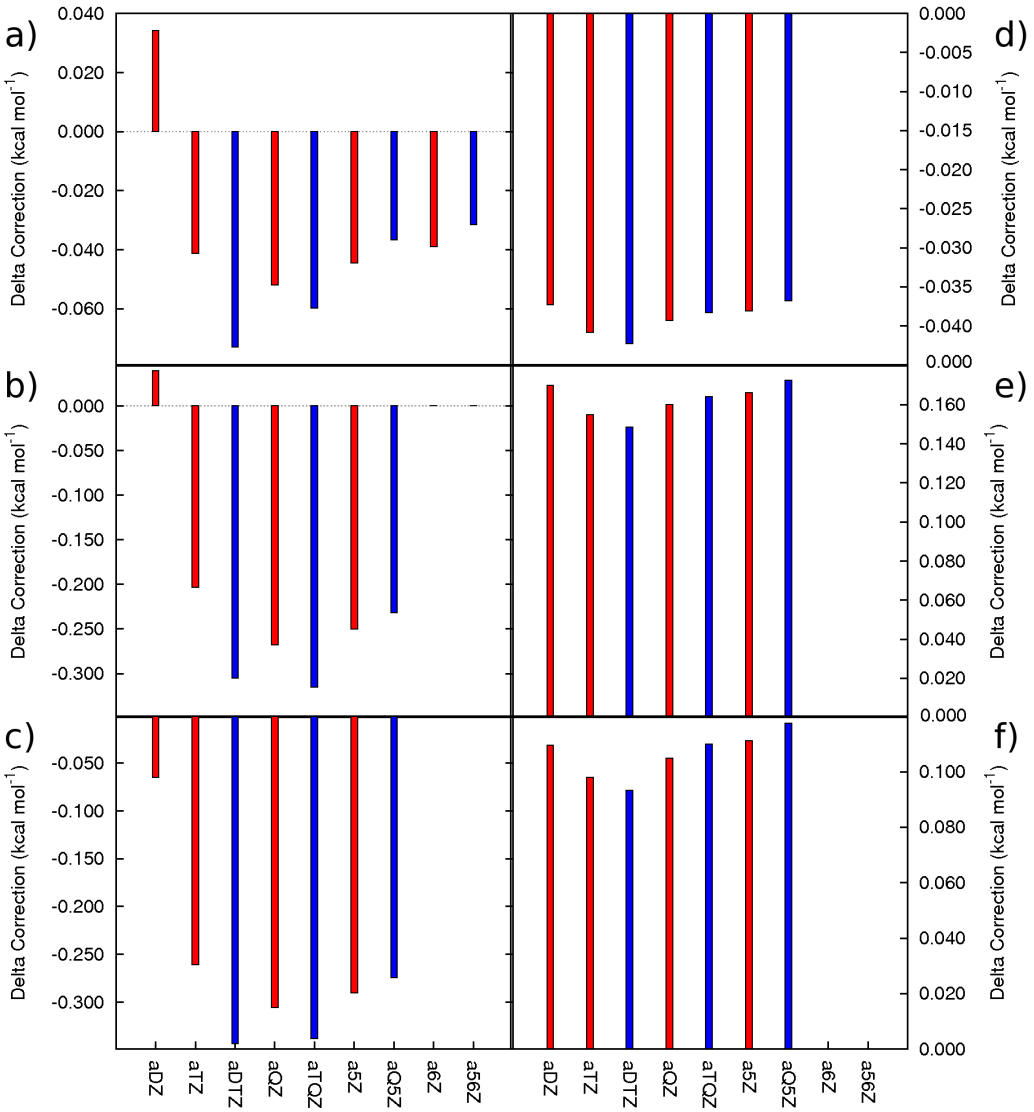

Figure 10: Basis set convergence of $\delta_{MP2}^{CCSD(T)}$ for a) water dimer, b) formic acid dimer, c) formamide dimer, d) methane dimer, and e) ethene·ethine complex, and f) ethene dimer. Hierarchical Dunning basis sets aDZ–a6Z (red bars) as well as their two-point Helgaker extrapolations (blue bars) are plotted, showing poor reliability of double-ζ results. All systems exhibit a "turnover" basis before which CBS estimates are unreliable and after which the term converges monotonically.

The $\delta_{MP2}^{CCSD(T)}$ correction for the water dimer in Figure 10a grows with increasing

basis set until aQZ, but decreases with the a5Z and a6Z basis sets. To achieve a nearly-converged CBS $\delta_{MP2}^{CCSD(T)}$ correction, one would have to acquire a CBS(aQZ,a5Z) extrapolated estimate. Because of the scaling of CCSD(T), employing basis sets of this size is infeasible for any but the smallest complexes. From Figure 10a, we also note how CBS(aDZ,aTZ) significantly overestimates the best available [a6Z or CBS(a5Z,a6Z)] estimates (by ∼0.04 kcal mol^{-1} out of a ∼0.03 kcal mol^{-1} correction). This is partly due to the poor quality of the aDZ basis set, but also due to the fact that $\delta_{MP2}^{CCSD(T)}$ terms do not converge monotonically with increasing basis set size.

We apply this same analysis to the double hydrogen-bonded system formic acid dimer, shown in Figure 10b. Again, we note that the $\delta_{MP2}^{CCSD(T)}$ correction does not converge monotonically and this causes the CBS(aDZ,aTZ) estimate to overshoot by -0.07 kcal mol^{-1} compared to CBS(aQZ,a5Z). Figure 10c demonstrates the same trends for formamide dimer. In this case, the aDZ basis provides the correct sign for $\delta_{MP2}^{CCSD(T)}$, but it is only -0.07 kcal mol^{-1} compared to a best estimate of -0.27 kcal mol^{-1}. The CBS(aDZ,aTZ) extrapolation overshoots the best estimate by 0.07 kcal mol^{-1}.

While these errors are not very large, they are undesirable and possible to avoid in general without costlier computations. Based on the water dimer, formic acid dimer, and formamide dimer test cases it appears that the aTZ basis set provides a nice Pauling point for the $\delta_{MP2}^{CCSD(T)}$ correction for hydrogen bonded systems. For two out of three cases, the aDZ basis provided the wrong sign for $\delta_{MP2}^{CCSD(T)}$, and in the other case, it achieved only one-fourth of the true value. Hence, for small hydrogen-bonded systems, we urge caution in using double-ζ basis sets for $\delta_{MP2}^{CCSD(T)}$ corrections for benchmark quality work. For somewhat larger complexes, contributions from dispersion forces will grow, and double-ζ basis sets (which work reasonably well for $\delta_{MP2}^{CCSD(T)}$ corrections in dispersion-dominated or mixed complexes) may perform better.

Methane dimer was investigated to see if similar qualitative basis set effects could

be seen in dispersion bound complexes. Figure 10d illustrates the $\delta_{MP2}^{CCSD(T)}$ correction for the Dunning basis set series for methane dimer. Again, we see non-monotonically converging $\delta_{MP2}^{CCSD(T)}$ correction, but now the magnitude of the difference between CBS(aDZ,aTZ) and either aQZ and CBS(aQZ,a5Z) is 0.003 and 0.005 kcal mol^{-1}, respectively. While such errors are negligible, CBS(aDZ,aTZ) is still not recommended in place of single-basis $\delta_{MP2}^{CCSD(T)}$ because of the consistent over-correction. Ethene·Ethine was investigated as a representative from the mixed category of the S22 test set. Figure 10e once again shows that the $\delta_{MP2}^{CCSD(T)}$ correction does not converge monotonically, instead featuring a turning point at aTZ. This leads CBS(aDZ,aTZ) extrapolation to overestimate the $\delta_{MP2}^{CCSD(T)}$ correction by 0.020 kcal mol^{-1} compared to the CBS(aQZ,a5Z) value. Again, these errors are not very large, but neither is the magnitude of the $\delta_{MP2}^{CCSD(T)}$ correction. As the size of the $\delta_{MP2}^{CCSD(T)}$ correction increases for other van der Waals dimers, this error should increase as well, and using aTZ for $\delta_{MP2}^{CCSD(T)}$ instead of CBS(aDZ,aTZ) may offer a more noticeable improvement.

3.4.2 Impact on Current Benchmark Sets for Non-Covalent Interactions

From this basis set study on the $\delta_{MP2}^{CCSD(T)}$ correction, we reach two general conclusions. 1) Extrapolated corrections should be avoided unless one can ensure that the basis sets used are beyond the turning point (typically aQZ for hydrogen-bonded and aTZ for mixed and dispersion bound complexes). These are admittedly large basis sets for CCSD(T) computations, which argues against using extrapolation techniques for $\delta_{MP2}^{CCSD(T)}$ corrections in general. 2) aDZ often results in the wrong sign for $\delta_{MP2}^{CCSD(T)}$ corrections for small hydrogen-bonded complexes (S22-1 through S22-7), and larger basis sets should be used whenever possible. In light of these new findings, we have decided to revise some existing benchmark sets that used extrapolated $\delta_{MP2}^{CCSD(T)}$ values. The best estimates for interaction energies of the S22 complexes have been revised as described in Table 3 and are herein referred to as S22B.

Table 3: Benchmark interaction energies (kcal mol^{-1}) for the S22B database with references from which each component was taken.

	Complex	Benchmark IE	Level of Theory
1	HB Ammonia Dimer, C_{2h}	-3.133	CCSD(T)/CBS(aQZ,a5Z)[4]
2	HB Water Dimer, C_s	-4.989	CCSD(T)/CBS(a5Z,a6Z)[4]
3	HB Formic Acid Dimer, C_{2h}	-18.753	CCSD(T)/CBS(aQZ,a5Z)[4]
4	HB Formamide Dimer, C_{2h}	-16.062	CCSD(T)/CBS(aQZ,a5Z)[4]
5	HB Hydrogen-Bonded Uracil Dimer, C_{2h}	-20.641	MP2/CBS(haQZ-ha5Z)[2] + $\delta_{MP2}^{CCSD(T)}$/aTZ+MB[1]
6	HB 2-Pyridone·2-Aminopyridine, C_1	-16.934	MP2/CBS(haQZ-ha5Z)[2] + $\delta_{MP2}^{CCSD(T)}$/aTZ+MB[1]
7	HB Adenine·Thymine WC, C_1	-16.660	MP2/CBS(haQZ-ha5Z)[2] + $\delta_{MP2}^{CCSD(T)}$/aTZ[3]
8	DD Methane Dimer, D_{3d}	-0.527	CCSD(T)/CBS(aQZ,a5Z)[4]
9	DD Ethene Dimer, D_{2d}	-1.472	CCSD(T)/CBS(aQZ,a5Z)[4]
10	DD Benzene·Methane, C_3	-1.448	MP2/CBS(aQZ-a5Z)[1] + $\delta_{MP2}^{CCSD(T)}$/aTZ+MB[1]
11	DD Parallel Displaced Benzene Dimer, C_{2h}	-2.654	MP2/CBS(aQZ-a5Z)[1] + $\delta_{MP2}^{CCSD(T)}$/aTZ+MB[1]
12	DD Pyrazine Dimer, C_s	-4.255	MP2/CBS(aQZ-a5Z)[1] + $\delta_{MP2}^{CCSD(T)}$/aTZ+MB[1]
13	MX Stacked Uracil Dimer, C_2	-9.805	MP2/CBS(haQZ-ha5Z)[2] + $\delta_{MP2}^{CCSD(T)}$/aTZ+MB[1]
14	DD Stacked Indole·Benzene, C_1	-4.524	MP2/CBS(aQZ-a5Z)[1] + $\delta_{MP2}^{CCSD(T)}$/aTZ+MB[1]
15	MX Stacked Adenine·Thymine, C_1	-11.730	MP2/CBS(haQZ-ha5Z)[2] + $\delta_{MP2}^{CCSD(T)}$/aTZ[3]
16	MX Ethene·Ethine, C_{2v}	-1.496	CCSD(T)/CBS(aQZ,a5Z)[4]
17	MX Benzene·Water, C_s	-3.275	MP2/CBS(aQZ-a5Z)[1] + $\delta_{MP2}^{CCSD(T)}$/aQZ[4]
18	MX Benzene·Ammonia, C_s	-2.312	MP2/CBS(aQZ-a5Z)[1] + $\delta_{MP2}^{CCSD(T)}$/aQZ[4]
19	MX Benzene·Hydrogen Cyanide, C_s	-4.541	MP2/CBS(aQZ-a5Z)[1] + $\delta_{MP2}^{CCSD(T)}$/aQZ[4]
20	DD T-shaped Benzene Dimer, C_{2v}	-2.717	MP2/CBS(aQZ-a5Z)[1] + $\delta_{MP2}^{CCSD(T)}$/aTZ+MB[1]
21	MX T-shaped Indole·Benzene, C_1	-5.627	MP2/CBS(aTZ-aQZ)[1] + $\delta_{MP2}^{CCSD(T)}$/aTZ+MB[1]
22	MX Phenol Dimer, C_1	-7.097	MP2/CBS(haQZ-ha5Z)[2] + $\delta_{MP2}^{CCSD(T)}$/aTZ+MB[1]

This new set removes any CBS extrapolated $\delta_{MP2}^{CCSD(T)}$ corrections and uses the largest underlying MP2/CBS as well as the largest $\delta_{MP2}^{CCSD(T)}$ correction available from the literature. We also provide larger basis set $\delta_{MP2}^{CCSD(T)}$ corrections or directly extrapolated CCSD(T)/CBS values (without using a $\delta_{MP2}^{CCSD(T)}$ term) for some of the smallest complexes. From Table 4, we see noticeable deviations for the hydrogen-bonded complexes compared to the S22A benchmark values of Takatani *et al.*[211] that used CBS(aDZ,aTZ) $\delta_{MP2}^{CCSD(T)}$ corrections.

Table 4: S22B benchmark interaction energies and differences (kcal mol^{-1}) with respect to Jurecka *et al.*, (Ref. [103], original publication), Podeszwa *et al.*, (Ref. [161], revision), and Takatani *et al.* (Ref. [211], S22A revision) literature values.

Complex	Jurecka	Podeszwa	Takatani	S22B	|ΔJurecka|	|ΔPodeszwa|	|ΔTakatani|
1	-3.17	-3.145	-3.150	-3.133	0.037	0.012	0.037
2	-5.02	-5.004	-5.070	-4.989	0.031	0.015	0.031
3	-18.61	-18.751	-18.810	-18.753	0.143	0.002	0.047
4	-15.96	-16.063	-16.110	-16.062	0.101	0.002	0.059
5	-20.65	-20.643	-20.690	-20.641	0.009	0.002	0.049
6	-16.71	-16.938	-17.000	-16.934	0.224	0.004	0.066
7	-16.37	-16.555	-16.740	-16.660	0.290	0.105	0.080
8	-0.53	-0.530	-0.530	-0.527	0.003	0.003	0.003
9	-1.51	-1.483	-1.480	-1.472	0.039	0.012	0.029
10	-1.5	-1.448	-1.450	-1.448	0.052	0.000	0.002
11	-2.73	-2.654	-2.620	-2.654	0.076	0.000	0.034
12	-4.42	-4.255	-4.200	-4.255	0.165	0.000	0.055
13	-10.12	-9.783	-9.740	-9.805	0.315	0.022	0.065
14	-5.22	-4.524	-4.590	-4.524	0.696	0.000	0.066
15	-12.23	-11.856	-11.660	-11.730	0.500	0.126	0.070
16	-1.53	-1.503	-1.500	-1.496	0.034	0.007	0.014
17	-3.28	-3.280	-3.290	-3.275	0.005	0.005	0.015
18	-2.35	-2.320	-2.320	-2.312	0.038	0.008	0.008
19	-4.46	-4.540	-4.550	-4.541	0.081	0.001	0.009
20	-2.74	-2.717	-2.710	-2.717	0.023	0.000	0.007
21	-5.73	-5.627	-5.620	-5.627	0.103	0.000	0.007
22	-7.05	-7.097	-7.090	-7.097	0.047	0.000	0.007
				Average	0.137	0.015	0.035
				Max	0.696	0.127	0.080

We also note substantial differences for the adenine·thymine complexes compared to the work of Podeszwa *et al.*[161], due to the latter's use of an aDZ $\delta_{MP2}^{CCSD(T)}$ correction. S22B benchmark values differ from those of Jurecka *et al.*,[103] Podeszwa *et al.*,[161] and Takatani *et al.*[211] by 0.137, 0.015, and 0.035 kcal mol^{-1} on average, respectively. Maximum differences are 0.696, 0.127, and 0.080 kcal mol $^{-1}$, respectively.

The NBC10[193, 213, 87] test set also generally utilized CBS(aDZ,aTZ) and CBS(haDZ,haTZ) $\delta_{MP2}^{CCSD(T)}$ corrections, so we updated these to use only the aTZ and haTZ $\delta_{MP2}^{CCSD(T)}$ corrections. The geometries were not changed. Supplemental Tables S2–S6 reflect the new values and the change to the new $\delta_{MP2}^{CCSD(T)}$ corrections. This new benchmark will be referred to as NBC10A. We report a shift of 0.017 kcal

mol^{-1} on average across all complexes and a maximum difference of 0.060 kcal mol^{-1} for PD Benzene Dimer at an intermolecular separation of 3.2 angstroms and a slip distance of 0.2 angstroms. We note the largest corrections are for stacked configurations on the repulsive wall.

Since the overestimation of $\delta_{MP2}^{CCSD(T)}$ corrections by CBS(aDZ,aTZ) extrapolation primarily affects hydrogen-bonded systems, we must also revise the HBC6 test set which consists of double- hydrogen-bonded complexes. The only modification to this test set was the replacement of CBS(aDZ,aTZ) $\delta_{MP2}^{CCSD(T)}$ corrections with pure aTZ $\delta_{MP2}^{CCSD(T)}$ corrections for all complexes. Tables 5-7 reflect the new values and the magnitude of the change to the new $\delta_{MP2}^{CCSD(T)}$ corrections, along with an incrementation of the test set name to "HBC6A". We report a shift of 0.073 kcal mol^{-1} on average across all complexes and a maximum difference of 0.257 kcal mol^{-1} for foramidine dimer at an intermolecular separation of 3.4 angstroms. Similar to NBC10, we note the largest corrections are for geometries on the repulsive wall.

Table 5: HBC6A benchmark interaction energies [using HF/aQZ, MP2/CBS(aTZ,aQZ) extrapolation, and $\delta_{MP2}^{CCSD(T)}$/aTZ components] and differences (kcal mol^{-1}) with respect to Thanthiriwatte *et al.* [Ref. [218], original publication using $\delta_{MP2}^{CCSD(T)}$/CBS(aDZ,aTZ) extrapolation]; Formic acid (FaOO) and formamide (FaON) dimers at various intermolecular distances (in Å).

Complex	**Original**	**HBC6A**	$\mid \Delta \mid$
FaOO–FaOO-3.4	-19.834	-19.627	0.207
FaOO–FaOO-3.5	-20.027	-19.850	0.177
FaOO–FaOO-3.6	-20.060	-19.910	0.150
FaOO–FaOO-3.7	-19.776	-19.650	0.126
FaOO–FaOO-3.8	-19.132	-19.027	0.105
FaOO–FaOO-3.9	-18.161	-18.075	0.086
FaOO–FaOO-4.0	-16.943	-16.873	0.070
FaOO–FaOO-4.1	-15.574	-15.517	0.057
FaOO–FaOO-4.2	-14.148	-14.100	0.048
FaOO–FaOO-4.3	-12.736	-12.697	0.039
FaOO–FaOO-4.4	-11.392	-11.360	0.032
FaOO–FaOO-4.6	-9.014	-8.990	0.024
FaOO–FaOO-4.8	-7.091	-7.074	0.017
FaOO–FaOO-5.0	-5.590	-5.577	0.013
FaOO–FaOO-5.4	-3.548	-3.539	0.009
FaOO–FaOO-5.8	-2.325	-2.323	0.002
FaOO–FaOO-6.4	-1.320	-1.320	0.000
FaOO–FaOO-7.0	-0.801	-0.802	0.001
FaOO–FaOO-8.0	-0.394	-0.397	0.003
FaOO–FaOO-10.0	-0.132	-0.135	0.003
FaON–FaON-3.4	-6.726	-6.556	0.170
FaON–FaON-3.5	-10.191	-10.027	0.164
FaON–FaON-3.6	-12.781	-12.628	0.153
FaON–FaON-3.7	-14.667	-14.529	0.138
FaON–FaON-3.8	-15.919	-15.796	0.123
FaON–FaON-3.9	-16.582	-16.475	0.107
FaON–FaON-4.0	-16.714	-16.622	0.092
FaON–FaON-4.1	-16.391	-16.313	0.078
FaON–FaON-4.2	-15.713	-15.647	0.066
FaON–FaON-4.3	-14.790	-14.735	0.055
FaON–FaON-4.4	-13.723	-13.678	0.045
FaON–FaON-4.6	-11.480	-11.448	0.032
FaON–FaON-4.8	-9.401	-9.379	0.022
FaON–FaON-5.0	-7.642	-7.626	0.016
FaON–FaON-5.4	-5.108	-5.097	0.011
FaON–FaON-5.8	-3.537	-3.528	0.009
FaON–FaON-6.4	-2.187	-2.181	0.006
FaON–FaON-7.0	-1.448	-1.443	0.005
FaON–FaON-8.0	-0.816	-0.813	0.003
FaON–FaON-10.0	-0.340	-0.337	0.003

Table 6: HBC6A benchmark interaction energies [using HF/aQZ, MP2/CBS(aTZ,aQZ) extrapolation, and $\delta_{MP2}^{CCSD(T)}$/aTZ components] and differences (kcal mol^{-1}) with respect to Thanthiriwatte *et al.* [Ref. [218], original publication using $\delta_{MP2}^{CCSD(T)}$/CBS(aDZ,aTZ) extrapolation]; dimers of formic acid (FaOO), formamide (FaON) and formamidine (FaNN) at various intermolecular distances (in Å).

Complex	**Original**	**HBC6A**	$\mid \Delta \mid$
FaNN–FaNN-3.4	-8.987	-8.730	0.257
FaNN–FaNN-3.5	-10.969	-10.725	0.244
FaNN–FaNN-3.6	-12.693	-12.463	0.230
FaNN–FaNN-3.7	-14.144	-13.932	0.212
FaNN–FaNN-3.8	-15.287	-15.106	0.181
FaNN–FaNN-3.9	-16.118	-15.950	0.168
FaNN–FaNN-4.0	-16.587	-16.440	0.147
FaNN–FaNN-4.1	-16.702	-16.575	0.127
FaNN–FaNN-4.2	-16.452	-16.344	0.108
FaNN–FaNN-4.3	-15.901	-15.811	0.090
FaNN–FaNN-4.4	-15.102	-15.028	0.074
FaNN–FaNN-4.6	-13.047	-12.999	0.048
FaNN–FaNN-4.8	-10.810	-10.780	0.030
FaNN–FaNN-5.0	-8.733	-8.715	0.018
FaNN–FaNN-5.4	-5.539	-5.532	0.007
FaNN–FaNN-5.8	-3.521	-3.517	0.004
FaNN–FaNN-6.4	-1.861	-1.861	0.000
FaNN–FaNN-7.0	-1.050	-1.051	0.001
FaNN–FaNN-8.0	-0.463	-0.466	0.003
FaNN–FaNN-10.0	-0.123	-0.127	0.004
FaOO–FaON-3.4	-14.356	-14.164	0.192
FaOO–FaON-3.5	-16.486	-16.312	0.174
FaOO–FaON-3.6	-17.833	-17.679	0.154
FaOO–FaON-3.7	-18.543	-18.409	0.134
FaOO–FaON-3.8	-18.692	-18.578	0.114
FaOO–FaON-3.9	-18.347	-18.250	0.097
FaOO–FaON-4.0	-17.592	-17.512	0.080
FaOO–FaON-4.1	-16.537	-16.471	0.066
FaOO–FaON-4.2	-15.300	-15.245	0.055
FaOO–FaON-4.3	-13.989	-13.944	0.045
FaOO–FaON-4.4	-12.684	-12.647	0.037
FaOO–FaON-4.6	-10.274	-10.248	0.026
FaOO–FaON-4.8	-8.245	-8.227	0.018
FaOO–FaON-5.0	-6.613	-6.597	0.016
FaOO–FaON-5.4	-4.330	-4.321	0.009
FaOO–FaON-5.8	-2.935	-2.931	0.004
FaOO–FaON-6.4	-1.753	-1.751	0.002
FaOO–FaON-7.0	-1.121	-1.119	0.002
FaOO–FaON-8.0	-0.598	-0.597	0.001
FaOO–FaON-10.0	-0.227	-0.228	0.001

Table 7: HBC6A benchmark interaction energies [using HF/aQZ, MP2/CBS(aTZ,aQZ) extrapolation, and $\delta_{MP2}^{CCSD(T)}$/aTZ components] and differences (kcal mol^{-1}) with respect to Thanthiriwatte *et al.* [Ref. [218], original publication using $\delta_{MP2}^{CCSD(T)}$/CBS(aDZ,aTZ) extrapolation]; dimers of formamidine (FaNN) with formamide (FaON) and formic acid (FaOO) at various intermolecular distances (in Å).

Complex	**Original**	**HBC6A**	$\mid \Delta \mid$
FaON–FaNN-3.4	-8.239	-8.021	0.218
FaON–FaNN-3.5	-10.918	-10.711	0.207
FaON–FaNN-3.6	-13.055	-12.862	0.193
FaON–FaNN-3.7	-14.717	-14.539	0.178
FaON–FaNN-3.8	-15.921	-15.763	0.158
FaON–FaNN-3.9	-16.672	-16.532	0.140
FaON–FaNN-4.0	-16.977	-16.856	0.121
FaON–FaNN-4.1	-16.865	-16.760	0.105
FaON–FaNN-4.2	-16.390	-16.301	0.089
FaON–FaNN-4.3	-15.631	-15.557	0.074
FaON–FaNN-4.4	-14.676	-14.614	0.062
FaON–FaNN-4.6	-12.490	-12.448	0.042
FaON–FaNN-4.8	-10.304	-10.277	0.027
FaON–FaNN-5.0	-8.362	-8.341	0.021
FaON–FaNN-5.4	-5.445	-5.434	0.011
FaON–FaNN-5.8	-3.617	-3.609	0.008
FaON–FaNN-6.4	-2.087	-2.082	0.005
FaON–FaNN-7.0	-1.295	-1.292	0.003
FaON–FaNN-8.0	-0.663	-0.661	0.002
FaON–FaNN-10.0	-0.237	-0.237	0.000
FaOO–FaNN-3.6	-26.289	-26.064	0.225
FaOO–FaNN-3.7	-24.035	-23.841	0.194
FaOO–FaNN-3.8	-23.017	-22.850	0.167
FaOO–FaNN-3.9	-22.133	-21.990	0.143
FaOO–FaNN-4.0	-21.122	-21.002	0.120
FaOO–FaNN-4.1	-19.920	-19.819	0.101
FaOO–FaNN-4.2	-18.544	-18.461	0.083
FaOO–FaNN-4.3	-17.056	-16.988	0.068
FaOO–FaNN-4.4	-15.526	-15.471	0.055
FaOO–FaNN-4.6	-12.583	-12.546	0.037
FaOO–FaNN-4.8	-10.031	-10.006	0.025
FaOO–FaNN-5.0	-7.960	-7.942	0.018
FaOO–FaNN-5.4	-5.069	-5.058	0.011
FaOO–FaNN-5.8	-3.336	-3.328	0.008
FaOO–FaNN-6.4	-1.906	-1.900	0.006
FaOO–FaNN-7.0	-1.170	-1.166	0.004
FaOO–FaNN-8.0	-0.587	-0.584	0.003
FaOO–FaNN-10.0	-0.202	-0.200	0.002

The last test set revised in this study is HSG,[50] which is a benchmark formed by dissecting the reaction site of a bound protein-drug complex (HIV-II protease-indinavir) into twenty-one pairs of chemical fragments that are not necessarily at equilibrium geometries. Again, here we only replace the CBS(haDZ,haTZ) $\delta_{MP2}^{CCSD(T)}$ correction with a haTZ $\delta_{MP2}^{CCSD(T)}$ correction. Table 8 reflects the new values and corresponding shifts for HSG-A. We report an average change of 0.027 kcal mol^{-1} and a maximum change of 0.102 kcal mol^{-1}. The small average change is due to there being only a few hydrogen-bonded complexes in the HSG test set.

Table 8: HSG-A benchmark interaction energies [using HF/aQZ, MP2/CBS(aTZ,aQZ) extrapolation, and $\delta_{MP2}^{CCSD(T)}$/haTZ components] and differences (kcal mol^{-1}) with respect to Faver *et al.* [Ref. [50] values, using $\delta_{MP2}^{CCSD(T)}$/CBS(haDZ,haTZ) extrapolation].

	Complex	**Original**	**HSG-A**	$\mid \Delta \mid$
1	ala29-big	-0.519	-0.518	0.001
2	ala128-small	-2.181	-2.283	0.102
3	arg8	-2.451	-2.478	0.027
4	ash26-asp125	-16.445	-16.526	0.081
5	asp129-big	-18.984	-19.076	0.092
6	asp130	-6.009	-5.998	0.011
7	gly28-big	-3.301	-3.308	0.007
8	gly50-ring-big	-0.554	-0.581	0.027
9	gly50-v1	-5.038	-5.066	0.028
10	gly127	-7.532	-7.509	0.023
11	gly148	-6.279	-6.274	0.005
12	ile48-big	0.305	0.302	0.003
13	ile147	-2.087	-2.103	0.016
14	ile150-big	-1.376	-1.378	0.002
15	ile184	-0.853	-0.856	0.003
16	leu23-big	-1.097	-1.100	0.003
17	pro181	-1.504	-1.534	0.030
18	val33-big	-0.473	-0.472	0.001
19	val83	-1.569	-1.598	0.029
20	val132	0.391	0.378	0.013
21	wat200	-9.486	-9.538	0.052
			Ave	0.027
			Max	0.102

The remaining errors in these benchmark test sets are anticipated to be the following: basis set incompleteness error (BSIE) of the MP2/CBS, BSIE of the $\delta_{MP2}^{CCSD(T)}$ correction, core-valence correction, and higher-order excitation corrections. To examine these sources of error, we look at each for the S22 benchmark set. By comparing MP2/CBS(aTZ,aQZ) to MP2/CBS(aQZ,a5Z), we estimate the BSIE of the MP2/CBS to have an average percent error of 0.10% (maximum 0.22% for PD Benzene dimer). We esimate the upper bound of the BSIE of the $\delta_{MP2}^{CCSD(T)}$ by comparing $\delta_{MP2}^{CCSD(T)}$/aDZ and $\delta_{MP2}^{CCSD(T)}$/aTZ. This approach reveals an average percent error of 0.60% (maximum of 3.01% for PD Benzene dimer). To understand the remaining BSIE better, we looked at both the counterpoise corrected and non-counterpoise corrected $\delta_{MP2}^{CCSD(T)}$ corrections (Figures 11 and 12).

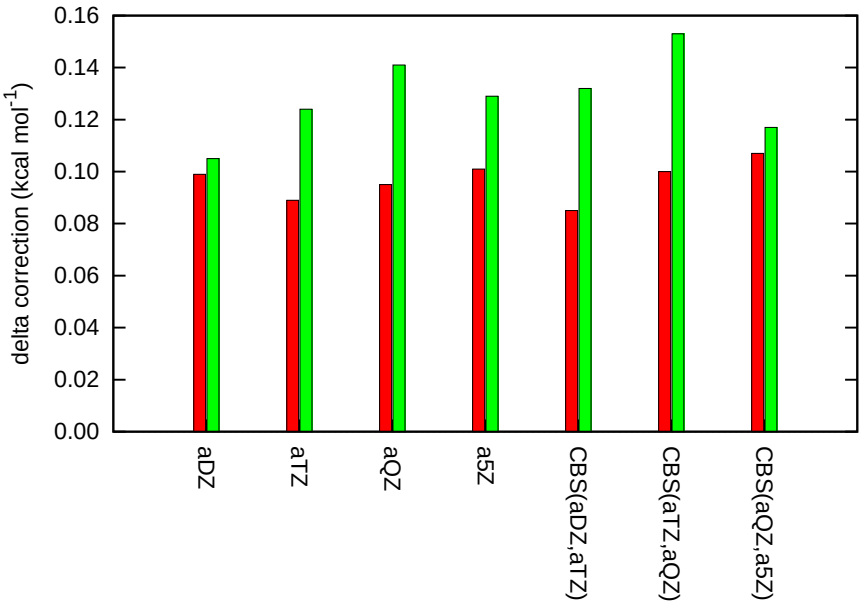

Figure 11: Ethene dimer (at the S22 geometry) counterpoise (red) and uncounterpoise (green) corrected $\delta_{MP2}^{CCSD(T)}$ corrections with aug-cc-pVXZ (X=D,T,Q,5). CBS values are computed with 2-point Helgaker extrapolation.

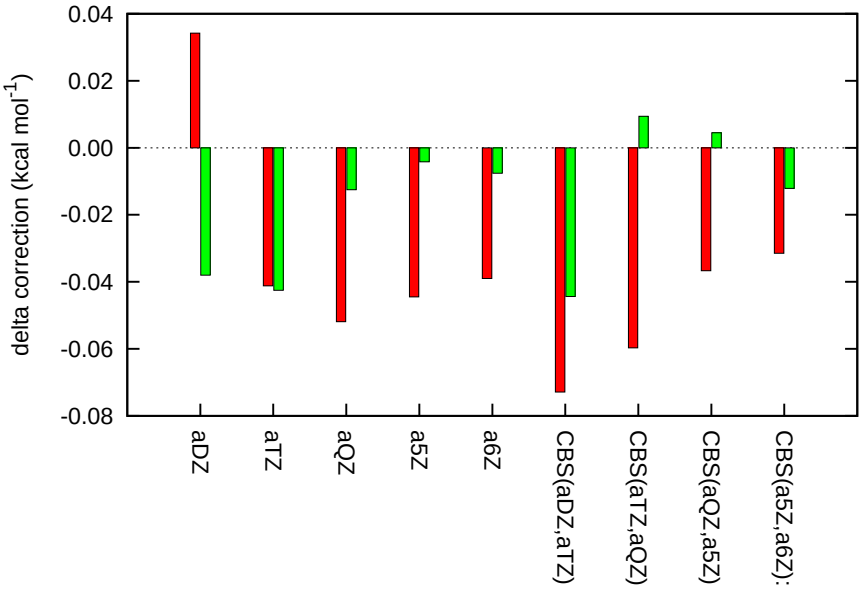

Figure 12: Water dimer (at the S22 geometry) counterpoise (red) and uncounterpoise (green) corrected $\delta_{MP2}^{CCSD(T)}$ corrections with aug-cc-pVXZ (X=D,T,Q,5,6). CBS values are computed with 2-point Helgaker extrapolation.

These figures reveal that there is still significant BSSE with the aTZ basis and that an uncorrected $\delta_{MP2}^{CCSD(T)}$ should therefore be used with caution (all our proposed benchmarks use the counterpoise-corrected $\delta_{MP2}^{CCSD(T)}$). The core-valence correction was estimated by Podeszwa *et al.*[161] to be on the order of 0.1% and no larger than 0.5% for all molecules in the S22 test set. Because of a lack of detailed studies of higher-order corrections (with an adequate basis set) in dispersion dominated complexes, post-triples corrections are hard to quantify. Hopkins *et al.*[93] estimated $\delta_{CCSD(T)}^{CCSD(TQ)}$ corrections to be approximately a tenth the magnitude of $\delta_{MP2}^{CCSD(T)}$ (with the same sign), but these were for relatively small systems. Pitonak *et al.*[157] reported quadruple excitation corrections for benzene dimer to be 0.04 kcal mol^{-1} (1.72%), but this study used a relatively small 6-31G*(0.25) basis set. A more detailed study of higher-order corrections with adequate basis sets is required before

giving bounds on this error.

3.4.3 Small Basis Set $\delta_{MP2}^{CCSD(T)}$ Corrections

Having established that benchmark-quality $\delta_{MP2}^{CCSD(T)}$ terms generally require triple-ζ basis sets to be truly robust, we examined the performance of small double-ζ basis sets often present in the literature. Table S1 in the supplemental material presents our best estimates of the $\delta_{MP2}^{CCSD(T)}$ correction for the S22 test set, along with estimates of this correction evaluated in various double-ζ basis sets. Mean absolute deviations are presented in Figure 13a.

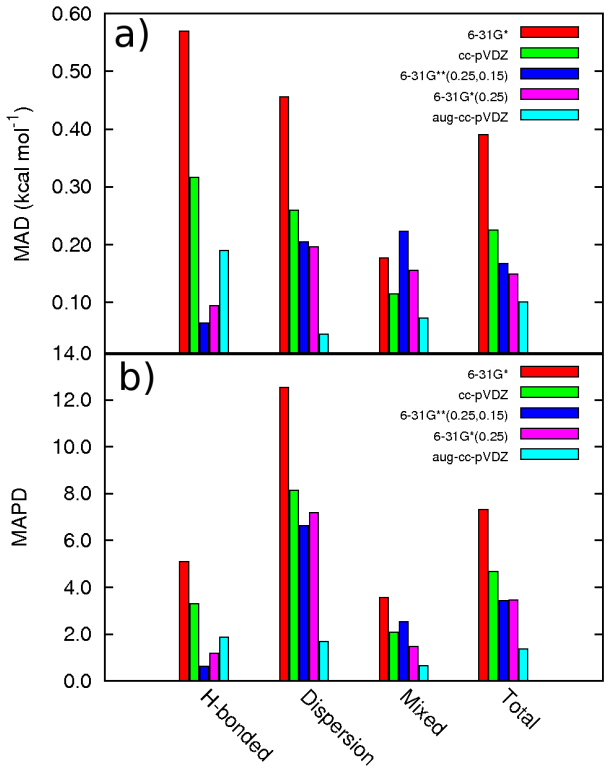

Figure 13: Performance of double-ζ basis sets for the $\delta_{MP2}^{CCSD(T)}$ correction. For the S22 test set, (a) mean absolute deviations and (b) mean absolute percent deviations are assessed in relation to S22B benchmark CCSD(T)/CBS values. While the specially-modified 6-31G*(0.25) and 6-31G**(0.25,0.15) basis sets yield low errors for hydrogen-bonding, they are significantly worse for dispersion-bound systems, and only aug-cc-pVDZ reaches an overall MAD for S22B of less than 0.1 kcal mol^{-1}.

It is clear that the original 6-31G* basis should not be used for these types of computations, as the optimized version 6-31G*(0.25) significantly outperforms it for the same computational cost. 6-31G*(0.25) performs relatively well for hydrogen-bonded complexes (MAD is 0.09 kcal mol^{-1}), but the error becomes somewhat larger than desirable for mixed complexes (0.15 kcal mol^{-1}) and dispersion-dominated complexes

(MAD 0.20 kcal mol^{-1} and MAPD 6.8%). Figure 13 also illustrates the importance of diffuse functions, as progressing from cc-pVDZ to aDZ reduces the MAD from 0.22 to 0.10 kcal mol^{-1}. Such augmented basis sets are especially important for dispersion bound complexes, reducing the MAD from 0.26 to 0.05 kcal mol^{-1}. Unfortunately, even aDZ is not an adequate basis set for high-quality $\delta_{MP2}^{CCSD(T)}$ corrections for all binding types; it produces a MAD of 0.18 kcal mol^{-1} for hydrogen-bonded systems. Overall, Figure 13 confirms that none of the small double-ζ basis sets can produce an acceptable level of error for benchmark-quality interaction energies across all binding types, although with judicious choices, some may be sufficient for application studies or narrowly defined benchmarking tasks.

3.4.4 Explicitly Correlated $\delta_{MP2}^{CCSD(T)}$ Corrections

We have shown that aDZ can lead to significant error in $\delta_{MP2}^{CCSD(T)}$ in some cases, so here we wanted to test how well this basis set could perform in an explicitly correlated framework. In Figure 14, the method labeled "MP2/CBS + δF12/aDZ" utilizes explicitly correlated $\delta_{MP2-F12}^{CCSD(T^*)-F12}$ corrections in an aDZ basis.

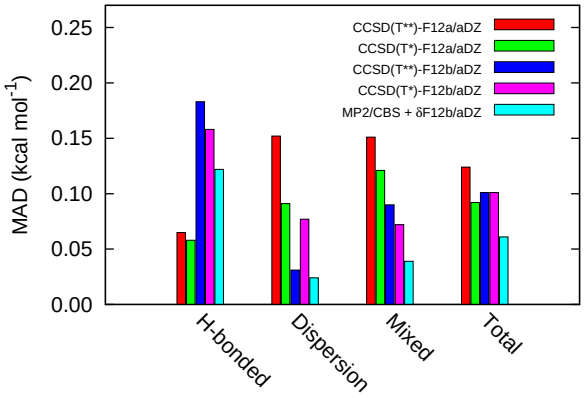

Figure 14: Mean absolute deviation (MAD) for various explicitly correlated methods using the aug-cc-pVDZ basis versus S22B benchmark CCSD(T)/CBS values. The method labeled "MP2/CBS + $\delta\,F12$b/aDZ" is included to show the best performance of an estimated CCSD(T) approach that uses a MP2/CBS and a $\delta_{MP2-F12}^{CCSD(T)-F12}$ correction in an aug-cc-pVDZ basis.

MP2/CBS + δF12b/aDZ achieves 0.06 kcal mol^{-1} MAD overall and 0.12, 0.02, and 0.04 for hydrogen-bonding, dispersion, and mixed bonding, respectively. While this method performs very well, it does not significantly improve upon the CCSD(T**)-F12b/aDZ approach itself. Apparently, the explicit correlation terms are so effective that CCSD(T**)-F12b/aDZ does not need to be mixed with MP2/CBS estimates to account for basis set effects. In particular, MP2/CBS + δF12/aDZ does not alleviate the maximum error incurred for hydrogen-bound complexes (0.23 kcal mol^{-1} for formic acid dimer). Hence, focal-point schemes using CCSD(T)-F12 to evaluate $\delta_{MP2-F12}^{CCSD(T^*)-F12}$ corrections do not seem to offer a large advantage over the underlying CCSD(T)-F12 in a modest basis.

3.5 Conclusions

This work examines the error incurred by employing polarized double-ζ basis sets for the $\delta_{MP2}^{CCSD(T)}$ portion of a focal-point estimate of CCSD(T)/CBS interaction energies for non-covalent complexes. The error in the coupled-cluster correction for a given basis set varies according to the non-covalent bonding motif. Although polarized double-ζ basis sets generally yield adequate estimates of $\delta_{MP2}^{CCSD(T)}$, especially if diffuse functions are included (or if the d exponent is made more diffuse), nevertheless the errors [MAD of 0.39, 0.15, 0.17, 0.23, 0.10 for the 6-31G*, 6-31G*(0.25), 6-31G**(0.25,0.15), cc-pVDZ, and aDZ for the S22 molecules] are too large for the resulting CCSD(T)/CBS values to be of true benchmark quality, given that several approximate methods are now capable of reproducing benchmark interaction energies within a few tenths of one kcal mol^{-1}. Hence, the remaining errors in $\delta_{MP2}^{CCSD(T)}$ should be taken into consideration when comparing new methods against benchmark sets that utilize $\delta_{MP2}^{CCSD(T)}$ corrections with such basis sets. By examining the coupled-cluster correction in progressively higher basis sets, up to a6Z, a characteristic turning point was found, after which the quantity converges monotonically and before which CBS extrapolations are unreliable. Particularly, CBS extrapolated $\delta_{MP2}^{CCSD(T)}$ corrections should not be used for hydrogen bonded complexes when employing aDZ and aTZ basis sets. We recommend simply using the largest single basis set affordable. We report revised benchmark values for the S22, NBC10, HBC6, and HSG test sets based on lessons learned in this work. Cartesian coordinates and revised interaction energies for these four test sets are available as supplementary material. The recent revision [1] of the S66 test set [175] used haDZ,haTZ extrapolated $\delta_{MP2}^{CCSD(T)}$ corrections; for those systems, just as for these, we expect the extrapolation procedure to introduce small errors (on the order of a few hundredths of one kcal mol^{-1}, perhaps more for any systems with double hydrogen bonds) relative to the true $\delta_{MP2}^{CCSD(T)}$/CBS values. We observe remarkable performance by the explicitly correlated methods

CCSD(T**)-F12a/b, even with a modest aDZ basis set, yielding an MAD of only 0.1 kcal mol^{-1} over the S22B test set. Such small errors mean that we need benchmark CCSD(T)/CBS values that are at least this precisely known, highlighting the value of the revised benchmark energies proposed here.

CHAPTER IV

AN ERROR AND EFFICIENCY ANALYSIS OF APPROXIMATIONS TO MØLLER–PLESSET PERTURBATION THEORY

4.1 Abstract

We present a systematic study of the synergistic effects of popular approximations to Møller–Plesset perturbation theory through second-order (MP2). This work applies the density-fitting (DF) approximation for two-electron integrals, the dual-basis (DB) approximation for the Hartree–Fock reference, and the use of "heavy-augmented" Dunning basis sets for basis set reduction, as well as combinations of these, to the S22 benchmark set of weakly bound dimers. For each approach, we report an error analysis as well as relative speed-ups for the 22 interaction energies in the set. Compared to the MP2/aug-cc-pVTZ level of theory, the DB-DF-MP2/heavy-aug-cc-pVTZ approach achieves an average speed-up of 18 with a root-mean-square error of only 0.076 kcal mol^{-1} (2%).

4.2 Introduction

In the last decade, the field of computational chemistry has demonstrated that high-level calculations on small molecules can in some cases achieve an accuracy comparable to that of experiment.[124, 166, 165, 135, 214] A current challenge lies in the development of approximations to robust levels of theory to address larger systems of interest. Dispersion-dominated interactions, for which dynamic electron correlation has been shown to play an important role,[221, 199, 197, 123, 158, 192] have attracted significant recent attention. When applying computational techniques to large-scale

problems, long-range interactions can accumulate and must be accounted for properly. While an accurate description is provided by highly-correlated methods such as coupled-cluster theory,[170, 190, 188] the steep computational cost of such methods constrains their applicability to systems of but a few dozen atoms with modest basis sets. To overcome this problem, ongoing research efforts focus on two fronts: (1) the modification of established methods by adding adjustable parameters fit to experiment or a higher-level of theory or (2) the development of approximations to robust levels of theory that maintain their inherent accuracy while reducing the cost.

The strategy of incorporating *ad hoc* terms with fitted parameters has shown great success for methods such as DFT-D (which adds a scaled, damped dispersion correction to a DFT functional).[240, 63, 64] However, there remains no means of systematically improving the accuracy, and such methods sometimes require numerous parameters trained upon specific test sets to produce high-quality results. Correlated wavefunction methods have also been modified by fitted parameters in spin-component-scaled Møller–Plesset perturbation theory (SCS-MP2),[62, 59, 79, 9, 213] spin-opposite-scaled Møller–Plesset perturbation theory (SOS-MP2),[99] and spin-component-scaled coupled-cluster with singles and doubles (SCS-CCSD).[212] MP2 tends to give reasonably reliable results for certain types of non-covalent interactions (such as alkane-alkane interactions and H-bonded interactions). In cases where MP2 exhibits significant errors (e.g., π-stacking interactions), the scaled MP2 methods such as SCS-MP2 tend to perform well.[9] Even in cases where very accurate binding energies are desired for non-covalent interactions, MP2 remains a critical ingredient in the theoretical procedure. When benchmark-quality results are needed, the current standard procedure is to evaluate the binding energies in the MP2 complete basis set limit, and then to correct for higher-order correlation terms by adding a ΔCCSD(T) correction [evaluated as the difference between CCSD(T) and MP2 binding energies in a smaller basis set].[221, 199] Thus, whether one uses bare MP2, scaled MP2, or

MP2 in conjunction with CCSD(T) corrections, MP2 computations remain important in studies of non-covalent interactions, and it is useful to explore approximations for speeding up these MP2 computations and to assess the associated errors.

In electronic structure theory, the evaluation and storage of four-index integrals is a common bottleneck. Various approaches to this problem have been explored, such as resolution of the identity[235, 46, 47, 54, 21, 108, 226, 233, 230] [now commonly referred to as density fitting (DF)], Cholesky decompositions[19, 180, 117, 12, 24, 227, 10, 246, 25, 33] (CD), and pseudo-spectral techniques.[137, 136, 56] In the DF treatment, four-index integrals $(\mu\nu \mid \rho\sigma)$ are approximated by summations over three-index quantities:

$$(\mu\nu \mid \rho\sigma) \approx \sum_{PQ} (\mu\nu \mid P)[J^{-1}]_{PQ}(Q \mid \rho\sigma), \tag{65}$$

where $[J^{-1}]_{PQ}$ is the inverse of the Coulomb metric evaluated in an auxiliary basis set,

$$[J]_{PQ} = \int P(\mathbf{r}_1)\frac{1}{\mathbf{r}_{12}}Q(\mathbf{r}_2)d^3\mathbf{r}_1 d^3\mathbf{r}_2. \tag{66}$$

The three-index quantity $(\mu\nu \mid P)$ serves to cast the product $(\mu\nu \mid$ onto the auxiliary basis via the Coulomb metric

$$(\mu\nu \mid P) = \int \mu(\mathbf{r}_1)\nu(\mathbf{r}_1)\frac{1}{\mathbf{r}_{12}}P(\mathbf{r}_2)d^3\mathbf{r}_1 d^3\mathbf{r}_2. \tag{67}$$

While density-fitting does not lower the asymptotic scaling of MP2, it does reduce the prefactor significantly, with speed-ups in the range of 2 to 5.5 reported.[233, 193] There exist many more methods for speeding up the evaluation of the correlation energy (e.g., local molecular orbital approaches such as local-MP2[233, 230]), yet the application of DF alone is often sufficient to reduce the cost of the correlation energy computation to the point that the time needed for the underlying self-consistent field (SCF) becomes the rate-determining step.

Numerous algorithmic advances have been achieved over the last three decades to improve SCF efficiency. These range from Pulay's direct inversion of iterative subspace[168, 169] (DIIS), which minimizes the number of SCF iterations, to modern linear scaling methods.[234, 207, 28, 186, 154] Two recent, similar advances in SCF theory are dual-basis techniques[238, 206, 204, 42, 203, 205] to project the SCF energy from a smaller basis set and perturbative corrections to estimate the SCF complete basis set limit.[41]

The dual-basis (DB) approximation proposed in the work of Steele *et al.* involves performing an iterative SCF in a small basis, and then taking a single Roothan diagonalization step in a larger target basis set. In practice, the small basis is typically a specially designed subset of the target basis set, although this restriction is not imposed by the theory. Once the SCF is converged with the small basis set, the occupied molecular orbital (MO) coefficients are projected onto the larger basis via

$$\mathbf{C}_{\overline{\mu}i} = \sum_{\overline{\mu\nu}} \sum_{\lambda} \mathbf{S}^{-1}_{\overline{\mu\nu}} \mathbf{S}_{\overline{\nu}\lambda} \mathbf{C}_{\lambda i}, \tag{68}$$

where $\mathbf{S}$ is the atomic orbital (AO) overlap matrix, i represents a MO index, Greek letters represent AO indices, and barred indices signify large-basis quantities. Using the newly constructed coefficient matrix, the new density matrix $\mathbf{P}$ is formed and a single Fock matrix is built and diagonalized. After including some first-order corrections, the DB-SCF energy is shown to be

$$E_{dual\ basis} = E_{small\ basis} + \sum_{\overline{\mu\nu}} \Delta\mathbf{P}_{\overline{\mu\nu}} \mathbf{F}_{\overline{\mu\nu}}, \tag{69}$$

where $\Delta\mathbf{P} = \mathbf{P}' - \mathbf{P}$ is the difference in the post-diagonalization density matrix $\mathbf{P}'$ and the small basis density matrix $\mathbf{P}$ (projected into the large basis). The small truncated basis sets used in the dual-basis methods have already been implemented[191] in the Q-Chem 3.2 program suite for several Pople and Dunning basis sets.

Another broadly employed approximation is the truncation of the aug-cc-pVXZ (X=D,T,Q) basis sets by eliminating diffuse functions from hydrogen atoms. These

truncated basis sets are commonly referred to as heavy-aug-cc-pVXZ and often abbreviated haXZ (X=D,T,Q). For biological applications and polymer studies, where a large number of hydrogens are present, haXZ can be a significant savings. Dropping augmented functions on hydrogen has been shown to have a small effect on properties such as interaction energies for non-bonded complexes.[193] The DF, DB, and haXZ approximations have all been developed independently. In this work, we systematically examine the practicability of combining these approximations and evaluate the magnitude of accumulated errors and attainable speed-ups. The S22 benchmark set[103] has been adopted because of its focus on non-covalent complexes, which are theoretically challenging.

Recent work by Steele *et al.* has shown that by combining DB and DF approximations within MP2, one can expect root mean square errors (RMSEs) of 0.043 and 0.019 kcal mol^{-1} for the MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ levels of theory, respectively, for the S22 set.[205] Their timings focus mainly on evaluating DB-DF-MP2 analytical gradients. In this work, we present an error and efficiency analysis for each approximation independently, then repeat with the S22 set for the combination of approximations, thereby permitting dissection of any errors incurred in the energy, as well as elucidating the origins of the speed-up. We also consider a series of linear alkanes to examine how these approximations behave as a function of increasing system size. Timings are compared to those from some other methods such as density functional theory.

4.3 Theoretical Approach

4.3.1 Efficiency Study of Approximate MP2 on Linear Alkanes

To evaluate gains by approximate MP2 methods, we examine a series of linear alkanes (C_nH_{2n+2}). Recent work has considered the effect of RI,[200, 201] Cholesky,[11, 245] and atomic-orbital based MP2[43] approximations on linear alkanes. Single-point

energy computations were performed using B3LYP, DB-B3LYP, MP2, DB-MP2, DF-MP2, and DB-DF-MP2 with the aug-cc-pVDZ basis set. For all dual-basis approximations, we employed the optimized basis sets of Steele *et al.*[205] referred to as racc-pVDZ, which has been shown to reproduce the target basis set (aug-cc-pVDZ) with minimal error in total energy. For the density-fitting auxiliary basis sets, we employed the optimized basis sets from Hättig and co-workers[228, 70] referred to as rimp2-aug-cc-pVDZ. In this work we will only be density fitting the MP2 contribution, not the underlying SCF, as that capability is not currently implemented in Q-Chem 3.2. The frozen-core approximation was employed for all the perturbative methods. Alkane geometries were constructed from the following parameters: $r_{CC} = 1.53$Å, $r_{CH} = 1.09$Å, and $\theta_{CCC} = 109.5^{o}$. For each level of theory, we report the overall user time as well as a decomposition of SCF and MP2 user times. For the tests performed, I/O time was typically minor, hence user times were very similar to wall times.

All computations were performed without taking advantage of spatial symmetry. All alkane computations used the Q-Chem 3.2[191] program suite on an Altus 1702 server featuring dual AMD Opteron 2378 processors (2.4 GHz, Quad Core), 32GB of DDR2 RAM, and 2x1TB 7200 RPM RAID-0 local disks. The SCF was converged to 10^{-8} hartree, and the integral threshold was 10^{-13}.

4.3.2 Performance Analysis of Approximate MP2 on the S22 Set

For a detailed analysis of the performance (both speed-ups and errors) by various approximate MP2 methods, we chose the S22 benchmark set,[103] which features diverse types of non-bonded interactions over a wide range of system sizes, from water dimer (6 atoms) to hydrogen-bonded adenine-thymine complex (30 atoms). Benchmark-quality CCSD(T)/CBS reference binding energies are available for this test set.[103, 211] For each of the 22 complexes, we report the interaction energy and the total user time for MP2, DB-MP2, DF-MP2, and DB-DF-MP2 with the

aug-cc-pVDZ, aug-cc-pVTZ, heavy-aug-cc-pVDZ, and heavy-aug-cc-pVTZ Dunning basis sets [a heavy-aug-cc-pVXZ (X=D,T) basis set consists of cc-pVXZ on hydrogen atoms and aug-cc-pVXZ on all other atoms]. The choice of the DB basis set and DF auxiliary basis is as described above. All interaction energies were corrected for basis-set superposition error (BSSE) using the counterpoise correction scheme outlined by Boys and Bernardi,[26] and individual calculations employed the frozen-core approximation. The benchmark machine for the S22 test set is an Intel Xeon (3.2 GHz, Single Core), 4GB of DDR2 RAM, and a 150 GB local disk. The SCF was converged to 10^{-8} hartree, and the integral threshold was 10^{-13}. This work focuses on approximating three computations: MP2/aug-cc-pVDZ, MP2/aug-cc-pVTZ, and MP2/CBS(aDZ,aTZ), where CBS(aDZ,aTZ) refers to a 2-point extrapolation as defined by Halkier *et al.*[67] using aug-cc-pVDZ and aug-cc-pVTZ correlation energies. All computations in this part of the study use one of these three canonical MP2 results as a reference point.

4.4 Results and Discussion

4.4.1 Efficiency of Approximations to MP2 for Linear Alkanes

In the following analysis, MP2 will be discussed in terms of two components: (a) the underlying SCF computation and (b) the evaluation of the MP2 correlation energy (including the transformation of the atomic orbital integrals to the molecular orbital basis). Using the alkane test cases, we investigate how the DF, DB, and haXZ approximations affect the speed and accuracy of the computation. For a medium-sized molecule such as $C_{20}H_{42}$, the underlying SCF takes 33% of the total user time as shown in Figure 15(a).

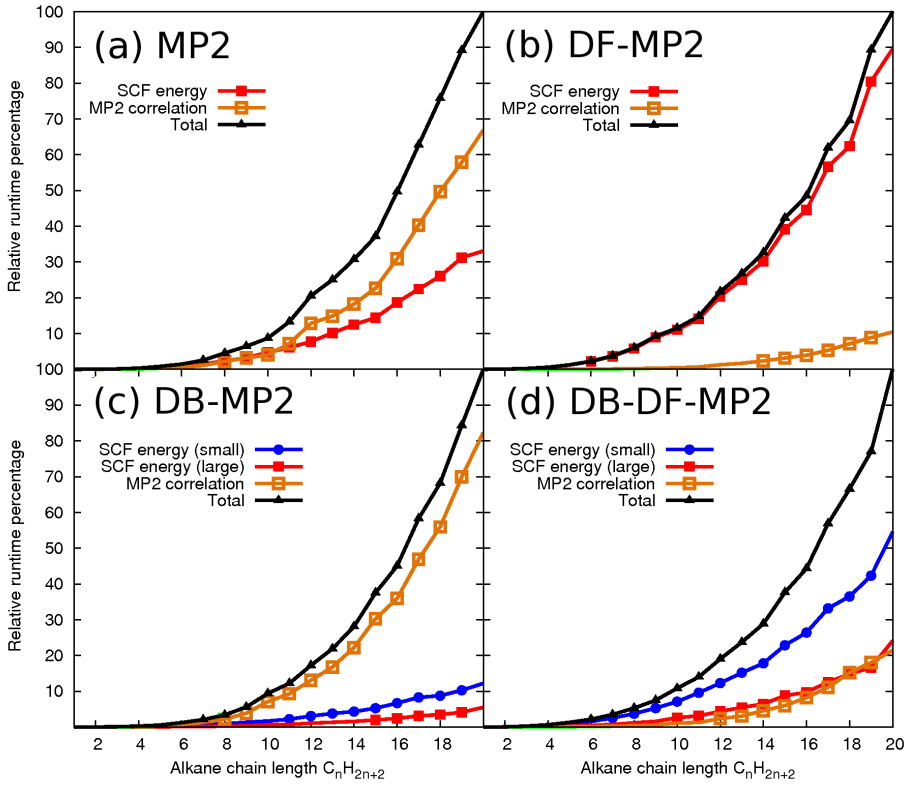

Figure 15: The relative user time and decompositions of (a) MP2 (b) DF-MP2 (c) DB-MP2 (d) DB-DF-MP2 are shown for the alkane series methane (CH_4) through dodecane ($C_{20}H_{42}$). The percentages are computed by taking the $C_{20}H_{42}$ as a reference, with the total broken into SCF and MP2 correlation components. For the DB approximations, SCF (small) refers to the percentage of time to solve the iterative part and SCF (large) refers to the percentage of time to perform the single Fock build in the target basis.

The MP2 contribution, which formally scales as $O(N^5)$, dominates over the SCF, which formally scales as $O(N^4)$, because of the AO to MO transformation required. (Of course, actual computational scalings with respect to system size will be lower

than these formal scalings because of techniques such as integral pre-screening; moreover, integral pre-screening will work best in one-dimensional systems such as these.) As system size or basis set increases, an increasing fraction of time will be spent in the MP2 portion of the computation. It is this costly step that DF abates by changing the complexity of the AO to MO transformation from $O(N^5)$ to $O(N^4)$ while increasing the correlation energy evaluation from $O(N^4)$ to $O(N^5)$, but with a much lower prefactor than before. Figure 15(b) shows how DF-MP2 reduces the time to compute the MP2 correlation energy, shifting the majority of the compute time to the underlying SCF. For $C_{20}H_{42}$, 90% of the time to compute the DF-MP2 energy is spent in the SCF, to obtain an overall speed-up of 2.46 relative to traditional MP2.

Now that the majority of the work has been shifted to the underlying SCF, we investigate dual-basis techniques that can drastically reduce the time to compute this stage. Figure 16 shows DB-MP2 speed-up to be only 1.18 relative to MP2, but this is only because of the large amount of time spent computing the correlation energy (and accordingly, the small cost of SCF) in the absence of the DF approximation.

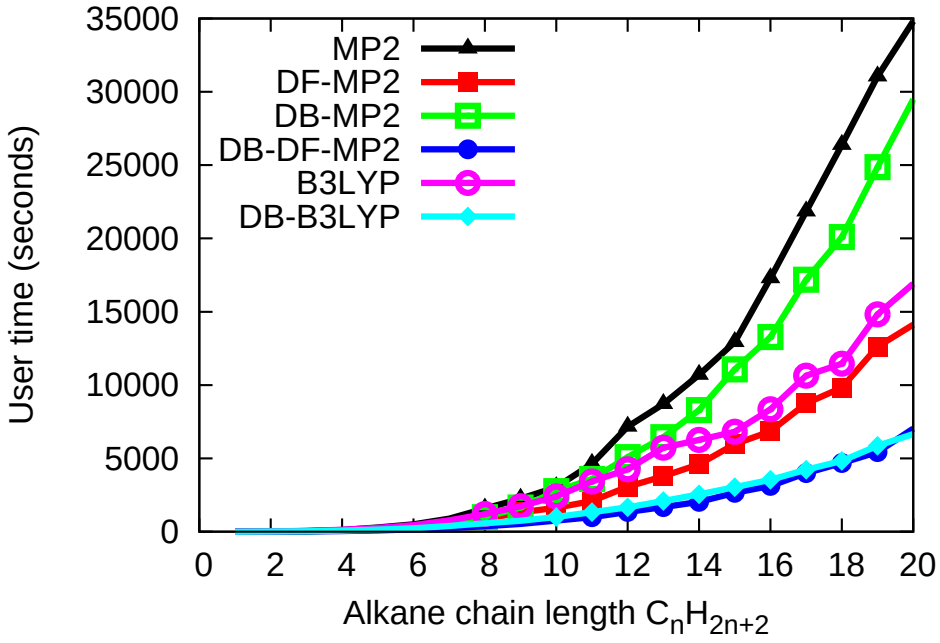

Figure 16: The total user times of MP2, DF-MP2, DB-MP2, DB-DF-MP2, B3LYP, and DB-B3LYP all with aug-cc-pVDZ basis set for the alkanes methane (CH_4) through dodecane ($C_{20}H_{42}$).

Figure 16 shows that the combination of the DF and DB approximations can yield speed-ups of 4.94 for $C_{20}H_{42}$. Within DB-DF-MP2, the bottleneck is the SCF iterations in the small basis, which consume 54% of the total compute time for $C_{20}H_{42}$. To put these improvements in context, we compared each method to B3LYP and dual-basis B3LYP (DB-B3LYP) in Figure 16, demonstrating that DF-MP2 is competitive with B3LYP and that DB-DF-MP2 is competitive with DB-B3LYP for system sizes upwards of $C_{20}H_{42}$. This implies that the overhead in computing the correlation contribution within DFT is comparable to the time to compute the density fitted correlation energy in MP2. Note, however, that DF-MP2 and DB-DF-MP2 are still not competitive with any pure DFT method that lacks Hartree–Fock exchange.

For the systems investigated, the errors of all of these approximations scale linearly

with system size. DB incurs an average error of 0.027 kcal mol^{-1} per atom, DF incurs an error of 0.006 kcal mol^{-1} per atom, and DB-DF incurs and error of 0.033 kcal mol^{-1} per atom.

4.4.2 Performance Analysis of Approximate MP2 on the S22 Set

To analyze the error introduced by the DB, DF, and haXZ approximations, we compute interaction energies for each of the 22 complexes in the S22 benchmark test set.[103] Table 9 presents the root-mean-square error (RMSE) for the test set at each level of theory.

Table 9: Mean unsigned error (MUE), root mean square error (RMSE), average percent error, and average speed-up analysis of approximating MP2/aug-cc-pVXZ and MP2/CBS(aDZ,aTZ) for the S22 test set of complexes. All errors in kcal mol^{-1}

Reference	Level of theory	Speedup	MUE	RMSE	%Error
MP2/aug-cc-pVDZ	MP2/haDZ	1.74	0.113	0.120	4.26
	DB-MP2/haDZ	2.61	0.106	0.120	4.05
	DF-MP2/haDZ	2.29	0.107	0.116	4.20
	DB-DF-MP2/haDZ	4.12	0.104	0.119	3.99
	DB-MP2/aDZ	1.78	0.034	0.043	0.82
	DF-MP2/aDZ	1.31	0.002	0.003	0.04
	DB-DF-MP2/aDZ	3.09	0.036	0.045[a]	0.84
MP2/aug-cc-pVTZ	MP2/haTZ	1.87	0.068	0.070	1.95
	DB-MP2/haTZ	5.76	0.072	0.077	2.01
	DF-MP2/haTZ	2.43	0.066	0.069	1.91
	DB-DF-MP2/haTZ	18.04	0.071	0.076	1.97
	DB-MP2/aTZ	3.25	0.012	0.017	0.20
	DF-MP2/aTZ	1.30	0.001	0.001	0.02
	DB-DF-MP2/aTZ	10.73	0.012	0.017[b]	0.19
MP2/CBS(aDZ,aTZ)	MP2/CBS(haDZ,haTZ)	1.86	0.039	0.044	0.94
	DB-MP2/CBS(haDZ,haTZ)	5.48	0.050	0.057	1.08
	DF-MP2/CBS(haDZ,haTZ)	2.42	0.038	0.043	0.92
	DB-DF-MP2/CBS(haDZ,haTZ)	15.73	0.049	0.056	1.05
	DB-MP2/CBS(aDZ,aTZ)	3.14	0.017	0.023	0.26
	DF-MP2/CBS(aDZ,aTZ)	1.30	0.001	0.001	0.01
	DB-DF-MP2/CBS(aDZ,aTZ)	9.65	0.017	0.022	0.26

[a] Reference [205] reports a RMSE of 0.043 kcal mol^{-1}. [b] Reference [205] reports a RMSE of 0.019 kcal mol^{-1}.

In trying to reproduce MP2/aug-cc-pVDZ results, we see that the RMSE from DF is 0.003 kcal mol^{-1} while achieving an average speed-up of 1.3. The magnitude of this error is very reasonable when compared to other remaining errors such as basis set incompleteness error (BSIE). On the other hand, the DB approximation incurs a RMSE of 0.043 kcal mol^{-1}. While an order of magnitude larger, the DB error is still rather small, especially considering it has a speed-up of 1.78. Applying both of the approximations simultaneously shows that they do indeed compound well, achieving a

speed-up of 3.1, but the errors are additive also, for a RMSE of 0.045 kcal mol^{-1}. The use of heavy-aug-cc-pVDZ in place of aug-cc-pVDZ yields a speed-up of 1.74 but at the cost of 0.120 kcal mol^{-1} average error. This average error is still small considering that the S22 MP2/aug-cc-pVDZ interaction energies range from -0.39 to -18.41 kcal mol^{-1}, but it may not be acceptable in some high-accuracy applications. The largest error introduced by neglecting diffuse functions on H atoms is 0.22 kcal mol^{-1} for the ethylene dimer test case, which has four closely packed hydrogens. Heavy-augmented basis sets should be avoided for systems with multiple hydrogen-hydrogen contacts, such as methane and ethene dimers.

To determine how well these approximations perform for larger basis sets, they were also tested against the MP2/aug-cc-pVTZ level of theory. For this larger basis set, RMSEs are reduced for all three approximations: DB has a RMSE of 0.017 kcal mol^{-1} (down from 0.043 kcal mol^{-1}), DF has a RMSE of 0.001 kcal mol^{-1} (down from 0.003 kcal mol^{-1}), and the use of heavy-augmented basis sets has a RMSE of 0.070 kcal mol^{-1} (down from 0.120 kcal mol^{-1}). When all three approximations are combined, speed-ups of 18.0 are achieved at the cost of 0.076 kcal mol^{-1} RMSE. Considering the large gain in computational efficiency, these errors are tolerable and DB-DF-MP2/haTZ is recommended for typical studies of non-bonded interactions.

We also examined how complete basis set (CBS) extrapolations affect the error for each approximation. CBS extrapolations consistently reduce the RMSE (shown at the bottom of Table 9) for the approximations considered. The extrapolations particularly abate the error caused by the use of heavy-augmented basis sets, reclaiming 0.026 kcal mol^{-1} on average. The compounding of all three approximations and CBS extrapolations [DB-DF-MP2/CBS(haDZ,haTZ)] yields a RMSE of 0.056 kcal mol^{-1} and a speed-up of 15.7. This speed-up is not quite as large as that observed for the DB-DF-MP2/haTZ (18.0), because the CBS extrapolations include haDZ computations which have a lesser efficiency gain. We note that MP2/CBS(aDZ,aTZ) has a

0.118 kcal mol^{-1} RMSE compared to MP2/CBS(aTZ,aQZ)[211] for the S22 test set.

The DF-MP2 speed-ups in our study are not as large as might be expected. We were forced us to use a core Hamiltonian guess to be consistent, because in Q-Chem 3.2, one cannot use the DB technique in conjunction with more advanced initial orbital guesses. The core Hamiltonian guess requires more SCF iterations to converge, thereby increasing the time spent in SCF. If superior SCF guesses were used, such as superposition of atomic densities (SAD) or a small basis projection, the overall computation would spend less time in the SCF and more time in the MP2 correlation. This would cause the DF methods to have better overall speed-ups and DB methods to have slightly smaller speed-ups.

To better understand the errors incurred through these approximations, a decomposition by binding type is shown in Figure 10.

Table 10: Mean unsigned error (MUE), root-mean-square error (RMSE), and average percent error for the interaction energies for each subgroup in the S22 test set in kcal mol^{-1}. All values are relative to MP2/CBS(aDZ,aTZ).

	H-bonding			Dispersion			Mixed		
Level of theory	MUE	RMSE	%Error	MUE	RMSE	%Error	MUE	RMSE	%Error
MP2/CBS(haDZ,haTZ)	0.059	0.061	0.64	0.029	0.033	1.26	0.031	0.032	0.88
DB-MP2/CBS(haDZ,haTZ)	0.057	0.065	0.56	0.051	0.057	1.60	0.043	0.047	1.00
DF-MP2/CBS(haDZ,haTZ)	0.058	0.060	0.63	0.027	0.032	1.21	0.030	0.031	0.86
DB-DF-MP2/CBS(haDZ,haTZ)	0.057	0.064	0.55	0.049	0.055	1.56	0.042	0.046	0.98
DB-MP2/CBS(aDZ,aTZ)	0.020	0.025	0.17	0.018	0.024	0.29	0.015	0.017	0.32
DF-MP2/CBS(aDZ,aTZ)	0.000	0.000	0.00	0.001	0.001	0.02	0.000	0.000	0.01
DB-DF-MP2/CBS(aDZ,aTZ)	0.019	0.025	0.17	0.018	0.023	0.28	0.015	0.017	0.32

Figure 3 and Ref. [103] define the division of complexes between hydrogen-bonded, dispersion-bound, and mixed-influence subgroups. As shown in Figure 10, dispersion-bound complexes experience a larger mean percent error than the hydrogen-bonded subset for every approximation examined, by a factor of 1.7–3.9, thereby suggesting

the the approximations examined in this work, particularly haXZ, may have difficulty with longer range interactions. For the CBS limit, we report errors among dispersion-dominated systems of 0.02%, 0.29%, 0.28%, and 1.26% for DF, DB, DB-DF, and heavy-augmented basis sets, respectively, while the corresponding value for hydrogen-bonded complexes in the last case is only 0.64%.

4.5 Conclusions

This work demonstrates that with a careful choice of approximations, MP2-quality results can be computationally affordable for systems with a few dozen atoms or larger without introducing significant error. Density fitting reduces the time to compute the MP2 correlation energy, dual-basis techniques abate the cost of the underlying SCF, and heavy-augmented functions speed up both parts of the computation relative to the fully augmented basis sets. Except for comparisons using the smaller heavy-aug-cc-pVDZ basis set, all of these approximations show significant speed-up while never incurring a RMSE greater than 0.045 kcal mol^{-1} for the S22 test cases. We also demonstrate that all of these approximations do indeed combine very efficiently. In future tests, density fitting will be extended to the SCF stage (currently not implemented in Q-Chem). The use of DF within the DB-SCF framework should be a significant stride toward achieving a level of theory that is not only accurate but applicable to a wide range of systems. Q-Chem also will soon have the capability to perform perturbative SCF approaches as outlined in the work of Gill *et al.*[41] These new computational tools will open up larger systems of interest to *ab initio* techniques while introducing errors which are negigible in most applications.

CHAPTER V

AVERAGING COUNTERPOISE- AND UNCOUNTERPOISE-CORRECTED INTERACTION ENERGIES AND $\delta_{MP2}^{CCSD(T)}$ CORRECTIONS FOR NON-BONDED COMPLEXES

Non-bonded interactions are critical to drug binding,[80, 126, 118, 51] biomolecular structures,[102, 143, 103, 176, 177], and crystal packing[178, 162, 163]. Accurate quantum modeling of the types of non-bonded interactions important to these applications, such as π-π or C-H/π interactions, requires the inclusion of dynamic electron correlation. One method that performs well is coupled-cluster theory with single, double, and perturbative triple excitations [CCSD(T)],[172] also known as the "gold standard" of computational chemistry. With sufficiently large basis sets, CCSD(T) has been shown to be very reliable for non-covalent interactions.[93] Because of the steep $\mathcal{O}(N^7)$ asymptotic scaling of its computational cost (where N represents the size of the system), CCSD(T) can only be applied to systems with a modest number of atoms (up to ∼30 atoms). To extend the range of applicability of large basis CCSD(T), a focal-point analysis[49, 37] approach is often used to estimate CCSD(T) in the large basis:

$$E_{CCSD(T)}^{large} \approx E_{MP2}^{large} + \delta_{MP2}^{CCSD(T)}, \tag{70}$$

$$\delta_{MP2}^{CCSD(T)} = E_{CCSD(T)}^{small} - E_{MP2}^{small}. \tag{71}$$

The subscripts and superscripts on energies refer to the method and basis set, respectively. The use of focal-point analysis has been applied to non-bonded complexes in

many recent studies.[221, 199, 101, 82, 101, 103, 198, 145, 23, 95, 155, 134, 51] While such a focal-point approach has been shown to be reliable, caution still needs to be taken when selecting the "small" basis set for the $\delta_{MP2}^{CCSD(T)}$ correction, as shown in recent work.[95, 134]

Counterpoise (CP) correction [26] is often used as a way to reduce basis set superposition error (BSSE) in computations of non-covalent interactions. Frequently, complexes tend to be overbound in uncorrected computations (because of BSSE), and underbound in CP-corrected computations. The converged result tends to be bounded by the CP-corrected and uncorrected (unCP) results, with convergence of the CP results being smoother and more systematic with respect to increasing the size of the basis set.[68] Hence, one pragmatic approach to computations of non-covalent interactions is to simply average CP and unCP values.[111, 110, 185, 109, 68, 112] For example, by considering a range of basis sets from aug-cc-pVDZ to aug-cc-pV5Z (abbreviated here as aDZ and a5Z, etc.), Halkier and co-workers[69, 68] have demonstrated that averaging CP and unCP results can yield accurate interaction energies in various small hydrogen-bonded complexes; they found that for aDZ basis sets, the unCP interaction energies are often closer the CBS limit than CP or averaged energies, but for basis sets aTZ and larger, averaging the CP and unCP interaction energies is a better approximation to the CBS limit. More recently, Mackie and DiLabio[130] have proposed to to extend the idea of averaging CP and unCP corrected interaction energies to focal-point methods as in Equation (71). In particular, they advocate adding averaged $\delta_{MP2}^{CCSD(T)}$ corrections to averaged MP2 interaction energies. Their approach is outlined as:

$$E_{Ave-CCSD(T)}^{large} \approx E_{Ave-MP2}^{large} + \delta_{Ave-MP2}^{Ave-CCSD(T)}, \tag{72}$$

$$\delta_{Ave-MP2}^{Ave-CCSD(T)} = E_{Ave-CCSD(T)}^{small} - E_{Ave-MP2}^{small}, \tag{73}$$

where any quantity with "Ave" is the average of CP and unCP corrected values. In

their work, and the present work, any CP corrections use the scheme outlined by Boys and Bernardi.[26]

Here we consider the S22 benchmark set,[103] which features diverse types of non-bonded interactions over a wide range of system sizes, from water dimer (6 atoms) to adenine·thymine complexes (30 atoms). The S22 set conveniently partitions complexes by binding type into hydrogen-bonding, dispersion-dominated, and mixed influence categories. This work uses the categorization provided by SAPT2+(3)/aTZ results.[91] Analysis of results by categories allows one to see if particular methods struggle for certain binding motifs. All reported errors in this work are relative to the S22B revised interaction energies.[134]

5.1 MP2 analysis

Before considering focal-point approaches, we first examine the MP2/CBS values and the $\delta_{MP2}^{CCSD(T)}$ corrections separately to see if averaging CP and unCP results is beneficial for either of these components. CP and unCP MP2 interaction energies as a function of basis set size (going from aDZ to a6Z) are presented for a few complexes in Figure 17.

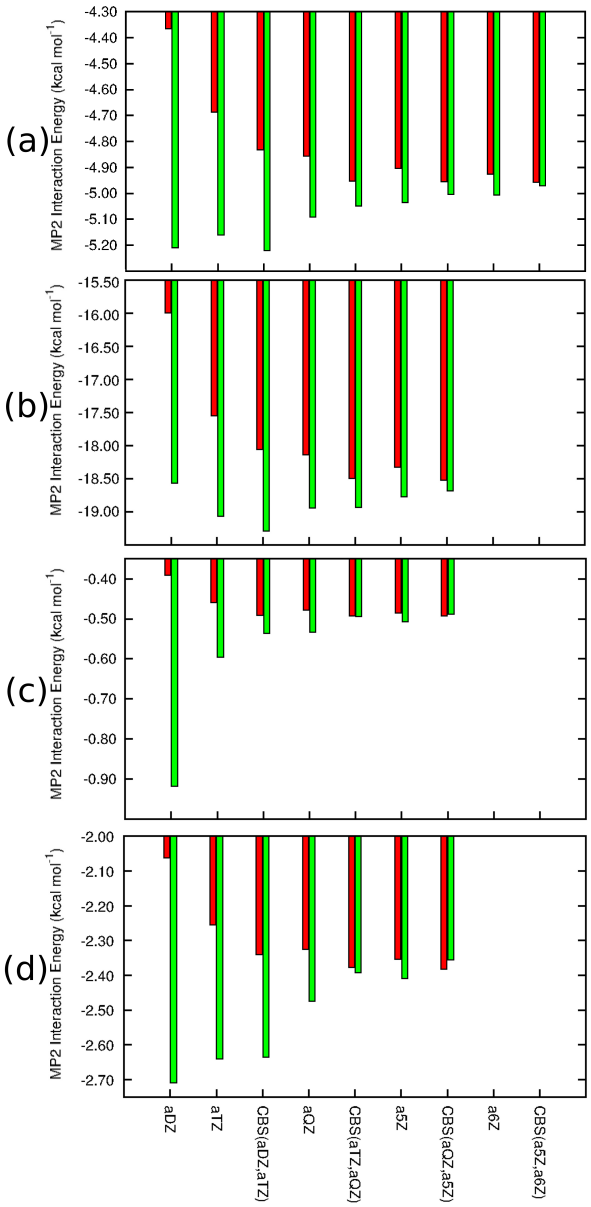

Figure 17: Counterpoise-corrected (red) and uncorrected (green) MP2 interaction energies with aug-cc-pVXZ (X=D,T,Q,5,6) for (a) water dimer (b) formic acid dimer (c) methane dimer and (d) sandwich cyanogen dimer. CBS values are computed with 2-point Helgaker extrapolation.

Water dimer from the S22 database,[103] shown in Figure 17(a), shows a rapid decrease of the BSSE with increasing basis set, going from 0.844 kcal mol^{-1} for the

aDZ basis, to 0.014 with CBS(a5Z,a6Z) extrapolation. Many benchmark studies of non-covalent interaction employing the focal-point approaches in equations 70 and 71 use an aQZ quality basis set or better for the MP2 component. With aQZ, the CP-corrected MP2 interaction energy of the water dimer is -4.856 kcal mol^{-1}, while the uncorrected value is -5.092 kcal mol^{-1}. Thus, even with an aQZ basis set, the BSSE is still 0.236 kcal mol^{-1}. Using the average of CP and unCP values, the MP2 interaction energy is -4.974 kcal mol^{-1}, which is significantly closer to the CBS(a5Z,a6Z) references of -4.957 (CP) and -4.971 (unCP) kcal mol^{-1} than are the individual components. MP2/CBS(aQZ,a5Z) extrapolation, which is the reference typically used in high-accuracy benchmarks such as the revised S22B test set,[134] reduces the BSSE significantly from its value with aQZ basis set, becoming only 0.049 kcal mol^{-1} for the water dimer. For such complete treatments of the basis set as this, the BSSE becomes so small that it makes little difference whether one uses CP, unCP, or averaged values: the results are -4.955, -5.000, and -4.980 kcal mol^{-1}, respectively. When employing basis sets as small as aDZ, the unCP interaction energy of -5.21 kcal mol^{-1} is significantly better than the CP interaction energy of -4.37 kcal mol^{-1} when compared to the CBS(a5Z,a6Z) extrapolation in the range of -4.957 (CP) to -4.971 (unCP) kcal mol^{-1}. Averaging in this small basis reduces the absolute error from 0.238 (unCP) to 0.183 kcal mol^{-1}. Employing the aTZ basis set, averaging -4.69 (CP) and -5.16 (unCP) kcal mol^{-1} yields and interaction energy of -4.924 kcal mol^{-1}, which is within 0.05 kcal mol^{-1} of the basis set limit, indicating that the averaging procedure can be beneficial once the aTZ basis set is reached. As for the aQZ results, averaging in an aTZ basis is better than using either of the components (CP or unCP) alone, at least for this test case.

To investigate doubly hydrogen-bonded complexes, we examine formic acid dimer from the HBC6 database.[218] Figure 17(b) shows that that BSSE is significantly larger than for single hydrogen bonded complexes like the water dimer, ranging from

2.574 kcal mol^{-1} with the aDZ basis set to 0.160 kcal mol^{-1} for CBS(aQZ,a5Z) extrapolation. While this is a large BSSE, it is only represents 0.9% of the interaction energy. Similarly to the water dimer, for the aQZ basis set, averaging CP (-18.057) and unCP (-19.294 kcal mol^{-1}) values results in an interaction energy of -18.675 kcal mol^{-1}, a better estimate of the CBS(aQZ,a5Z) limit [ranging from -18.521 (CP) to -18.682 (unCP) kcal mol^{-1}] than either component alone. When employing basis sets as small as aDZ, the unCP interaction energy of -18.564 kcal mol^{-1} is significantly better than the CP interaction energy of -15.990 kcal mol^{-1}, or the average value of -17.277 kcal mol^{-1}. For the aTZ basis set, averaging -17.552 (CP) and -19.068 (unCP) kcal mol^{-1} yields and interaction energy of -18.310 kcal mol^{-1}, which has errors of 0.161 (compared to CP/CBS limit) or 0.372 kcal mol^{-1} (compared to unCP/CBS limit). In this basis, as for the aQZ basis, the average appears to be closer to the CBS limit than either the CP value (error of 0.969 kcal mol^{-1} vs CP/CBS) or the unCP value (error of -0.386 kcal mol^{-1} vs unCP/CBS).

Hence, for both water dimer and formic acid dimer, averaging CP and unCP values seems to be an improvement over either value alone when either the aTZ or aQZ basis sets are used. Using CBS(aTZ,aQZ), CP corrected interaction energies perform best for both of these hydrogen bonded complexes. For even more complete treatments of basis set effects, such as CBS(aQZ,a5Z) extrapolation, little difference is seen between the results of CP, unCP, or averaging. On the other hand, for aDZ basis sets, unCP results are closer to the CBS limit for these two hydrogen-bonded test cases than are either CP or averaged results.

To repeat the same analysis for dispersion-dominated complexes, methane dimer from the S22 database,[103] and sandwich cyanogen dimer[92] were chosen. Methane dimer, shown in Figure 17(c), shows the BSSE of MP2/CBS(aQZ,a5Z) to be a mere 0.004 kcal mol^{-1} (0.9% of the interaction energy). The MP2/aQZ interaction energies are -0.478 (CP) and -0.533 (unCP) kcal mol^{-1}, while the average is -0.506 kcal mol^{-1}.

Again, averaging is closer the CBS(aQZ,a5Z) limit, which is in the range -0.492 (CP) to -0.488 (unCP) kcal mol^{-1}. When employing the aDZ basis set, the CP interaction energy of -0.390 is closer than the unCP interaction energy of -0.918 kcal mol^{-1}. Averaging these interaction energies yields an error of 0.162 kcal mol^{-1}, which is not an improvement over the CP estimate which has an error of 0.102 kcal mol^{-1} compared to the CP CBS limit. If aTZ were used, the unCP results is shifted significantly to -0.596 kcal mol^{-1}, while the CP changes very little with an interaction energy of -0.459 kcal mol^{-1}. Averaging yields a absolute deviation of 0.035 kcal mol^{-1}, where CP yields an error of 0.034 kcal mol^{-1}. Hence, for this dispersion-dominated case, averaging is beneficial in an aQZ basis, it is about the same as the CP result in an aTZ basis, and it is inferior to the straight CP result in an aDZ basis. This is similar to the pattern observed for the hydrogen bonded test cases, except that for the smallest (aDZ) basis set, now the CP result is superior to the unCP result.

Cyanogen dimer, shown in Figure 17(d), exhibits a BSSE of 0.028 kcal mol^{-1} for CBS(aQZ,a5Z) (1.2% of the interaction energy). MP2/aQZ interaction energies are -2.341 (CP) and -2.634 (unCP) kcal mol^{-1}, with an average of -2.488 kcal mol^{-1}. MP2/CBS(aQZ,a5Z) interaction energies are in the range of -2.383 (CP) to -2.355 (unCP) kcal mol^{-1}, and thus CP outperforms both average and unCP interaction energies in the aQZ basis. When using a smaller basis set such as aDZ, we find interaction energies of -2.063 (CP), -2.710 (unCP), and -2.386 (Ave) kcal mol^{-1} and see that averaging achieves an absolute error of less than 0.005 kcal mol^{-1}, which is good compared to the absolute errors of 0.320 (CP) and 0.327 (unCP) kcal mol^{-1}. Using the aTZ basis set, we report interaction energies of -2.255 (CP), -2.641 (unCP), and -2.448 (Ave) kcal mol^{-1}, where averaging performs best with an absolute error of 0.065 kcal mol^{-1}, compared to the absolute errors of 0.128 (CP) and 0.258 (unCP) kcal mol^{-1}.

This confirms the notion that averaging CP and unCP interaction energies can

provide a reliable estimate of the complete basis set limit. While averaging CP and unCP is useful when employing smaller basis sets, it is often not very beneficial at larger basis sets that are routinely used in benchmark computations because the remaining BSSE has already been reduced to around 1% of the interaction energy.

5.2 $\delta_{MP2}^{CCSD(T)}$ analysis

To investigate averaging CP and unCP $\delta_{MP2}^{CCSD(T)}$ values, Figure 18 plots $\delta_{CP-MP2}^{CP-CCSD(T)}$ and $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ values as a function of basis set size, going from aDZ to a6Z; also plotted are the complete basis set (CBS) extrapolations for each basis set pair considered using a 2-point Helgaker extrapolation.[67] Results are shown for the water dimer, the formic acid dimer, the sandwich cyanogen dimer, and the sandwich benzene dimer.

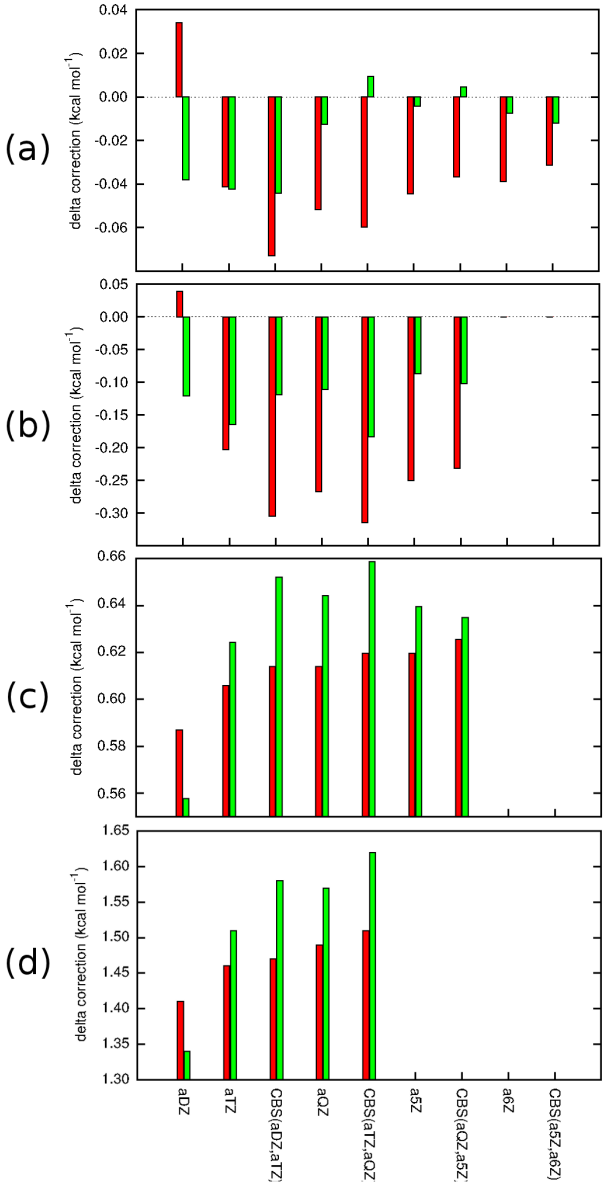

Figure 18: Counterpoise-correct (red) and uncorrected (green) $\delta_{MP2}^{CCSD(T)}$ corrections with aug-cc-pVXZ (X=D,T,Q,5,6) basis sets for (a) water dimer (b) formic acid dimer (c) sandwich cyanogen dimer and (d) sandwich benzene dimer. CBS values are computed with 2-point Helgaker extrapolation.

Let us begin with the water dimer results, shown in Figure 18(a). For the aDZ basis, $\delta_{CP-MP2}^{CP-CCSD(T)}$ fails to even predict the correct sign (208% error) whereas the $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ value matches the CBS estimate within 0.007 kcal mol^{-1}. Averaging

CP and unCP yields a $\delta_{CP-MP2}^{CP-CCSD(T)}$ correction of -0.002 kcal mol^{-1}, which is significantly worse than the unCP estimate. As the basis set is increased, however, the $\delta_{CP-MP2}^{CP-CCSD(T)}$ value is much more well behaved than the $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ variant. Indeed, the extrapolated $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ values show no reliable convergence behavior, in agreement with previous observations about the unsystematic convergence of electron correlation contributions to interaction energies of hydrogen-bonded systems.[111, 68] On the other hand, two-point extrapolated $\delta_{CP-MP2}^{CP-CCSD(T)}$ values show a very nice convergence pattern. The notion of unCP corrected interaction energies being more reliable for hydrogen-bonded complexes has been offered in recent studies,[187, 5] and in this example it is certainly true that the uncorrected $\delta_{MP2}^{CCSD(T)}$ values exhibit less error compared to the CBS limit, even though the values behave erratically with respect to basis set.

Figure 18(b) shows a similar trend for formic acid dimer. This test case shows that even when using CBS(aQZ,a5Z) extrapolation, there can remain considerable uncertainty in the CBS limit for the $\delta_{MP2}^{CCSD(T)}$ correction (the BSSE here is 0.130 kcal mol^{-1}). At least based on the results for water dimer and formic acid dimer, it would appear that for hydrogen-bonded systems, in an aDZ basis set uncorrected $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ values are the superior choice to CP-corrected or averaged values when using a focal-point approach like that in equations 70 and 71. For larger basis sets, both CP-corrected and uncorrected values for $\delta_{MP2}^{CCSD(T)}$ behave somewhat erratically with respect to basis set, although the CP values behave more systematically than unCP values in the case of water dimer.

Figures 18 (c) and 18 (d) present the CP and unCP corrected $\delta_{MP2}^{CCSD(T)}$ values as a function of basis set size for dispersion-dominated complexes. The cyanogen dimer (sandwich configuration)[92] shows relatively good agreement at the basis set limit between the CP and unCP corrected $\delta_{MP2}^{CCSD(T)}$ values, with results differing by only 0.009 kcal mol^{-1}. For this complex, using the aDZ basis set, CP corrections are

closer to the CBS limit, but for aTZ, the unCP result outperforms the CP result. CBS(aDZ,aTZ) extrapolation hurts the performance of the unCP corrected $\delta_{MP2}^{CCSD(T)}$ values with a maximum error of 0.028 kcal mol^{-1} [relative to our best estimate at CBS(aQZ,a5Z)]. For this complex, both CP and unCP are able to converge to the $\delta_{MP2}^{CCSD(T)}$ limit within 0.07 kcal mol^{-1} for all basis sets. Averaging is the best approach for the following basis sets: CBS(aDZ,aTZ), aQZ, and a5Z. UnCP performs best for aTZ, and CP only outperforms for the aDZ basis set and CBS(aTZ,aQZ) extrapolation.

For the benzene dimer (sandwich configuration, geometry from [58] and $\delta_{MP2}^{CCSD(T)}$ corrections taken from [193]) on the other hand, it is significantly harder to eliminate the BSSE. Figure 18d) shows that even for CBS(aTZ,aQZ) extrapolations, there still remains 0.11 kcal mol^{-1} of BSSE. While we cannot say definitively what the converged $\delta_{MP2}^{CCSD(T)}$ value is for this case, we can note that the CP corrected $\delta_{MP2}^{CCSD(T)}$ values appear to be converging much more smoothly (as we have seen in the four non-bonded complexes presented here).

Although this term is reputed to converge very quickly with respect to basis set size, we see that this is only true within certain error bars for large aromatic complexes.

To summarize the results from this section, when using an aDZ basis set to compute a coupled-cluster correction, $\delta_{MP2}^{CCSD(T)}$, to an interaction energy according to a focal-point approach such as that in Equation (70), unCP values appear to be the best choice for hydrogen-bound complexes, while CP values appear to be the best for dispersion-bound and mixed binding motifs. In the aDZ basis set, the averages of CP and unCP values for $\delta_{MP2}^{CCSD(T)}$ did not consistently outperform either unCP or CP for the test cases considered here. When using an aTZ basis set, $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ appears to provide the best estimate across all binding motifs when comparing against the CBS values. While unCP outperforms CP and averaging in the aTZ basis set, the

possible errors introduced remain relatively small (0.028 and 0.058 kcal mol^{-1} across the S10 and S22, respectively), but are expected to increase with system size.

5.3 Composite approach analysis

In Figure 19, we present errors versus the S22B benchmarks for various composite approaches that average the MP2/CBS and/or the $\delta_{MP2}^{CCSD(T)}$ employing the aDZ basis set.

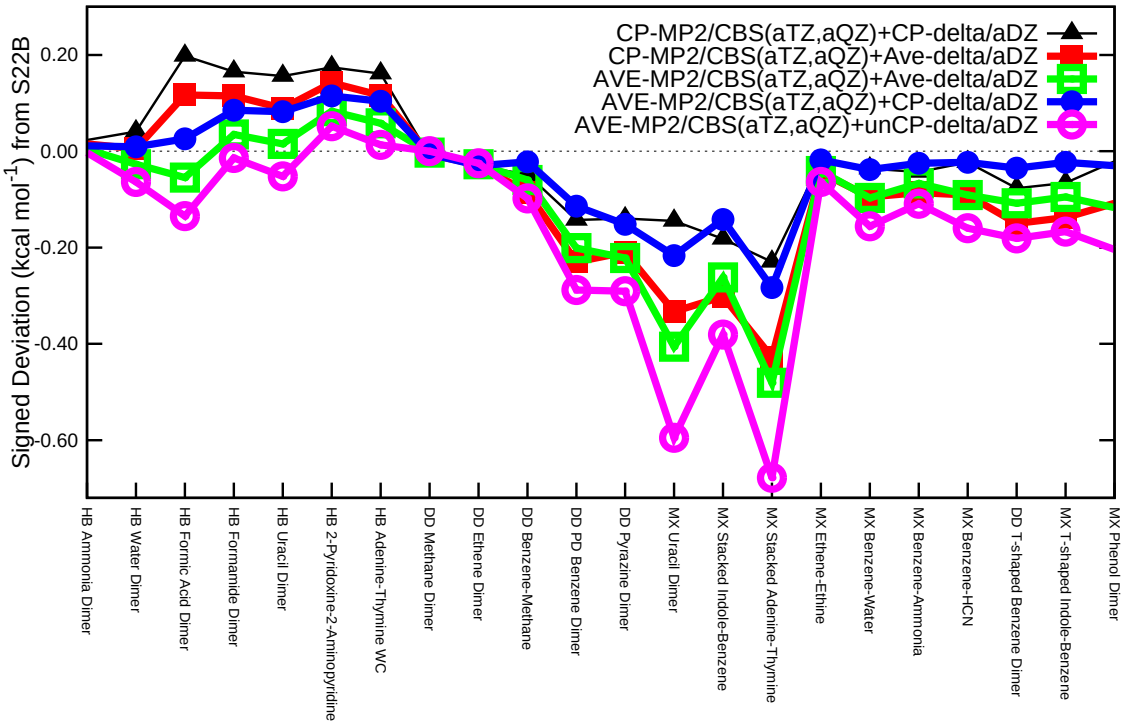

Figure 19: Signed errors of interaction energies versus S22B benchmark CCSD(T)/CBS values (Ref. [134]) for composite approaches using aug-cc-pVDZ.

CP-MP2/CBS(aTZ,aQZ) + $\delta_{CP-MP2}^{CP-CCSD(T)}$ is included as common method for comparison and has a mean absolute error (MAE) of 0.131, 0.099, 0.073, and 0.097 kcal mol^{-1} for hydrogen-bonded, dispersion-dominated, mixed influence, and overall categories, respectively (this ordering will be used for the rest of the analysis). Errors

are computed against the current best estimates from the S22B database.[134] Ave-MP2/CBS(aTZ,aQZ) + $\delta_{CP-MP2}^{CP-CCSD(T)}$ has MAEs of (0.062, 0.075, 0.082, and 0.072 kcal mol^{-1}). Thus, averaging the CP and unCP MP2/CBS(aTZ,aQZ) values reduces the MAE by a factor of two for hydrogen-bonded complexes, while not adding significant error to the description of dispersion-dominated complexes. Next, if we average the CP and unCP values for both the MP2 contribution and the $\delta_{MP2}^{CCSD(T)}$ component, Ave-MP2/CBS(aTZ,aQZ) + $\delta_{Ave-MP2}^{Ave-CCSD(T)}$ has MAEs of (0.040, 0.118, 0.174, and 0.116 kcal mol^{-1}). While this approach is able to describe hydrogen-bonded complexes extremely well, the error incurred for dispersion-dominated complexes (0.174 kcal mol^{-1}) is larger than the non-averaged approach (0.099 kcal mol^{-1}). Ave-MP2/CBS(aTZ,aQZ) + $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ has MAEs of (0.048, 0.160, 0.267, and 0.170 kcal mol^{-1}). This composite approach has increased error for both mixed influence and dispersion-dominated complexes relative to the non-averaged approach. Lastly, CP-MP2/CBS(aTZ,aQZ) + $\delta_{Ave-MP2}^{Ave-CCSD(T)}$ has MAEs of (0.0857, 0.141, 0.165, and 0.133 kcal mol^{-1}). Overall, from these errors, we can conclude that averaging CP and unCP MP2/CBS component of the focal-point method appears to best describe the S22 test set. When it comes to the question of how to best correct for higher order correlation, it is not as clear. For hydrogen-bonded cases, both $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ and $\delta_{ave-MP2}^{ave-CCSD(T)}$ are very accurate, but for dispersion bound and mixed influence cases, $\delta_{CP-MP2}^{CP-CCSD(T)}$ are the best choice. $\delta_{CP-MP2}^{CP-CCSD(T)}$ performing best for dispersion bound complexes is in good agreement with out earlier analysis of prototype systems, but $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ is not always better than $\delta_{ave-MP2}^{ave-CCSD(T)}$ for hydrogen bound complexes, which demonstrates some extra error cancellation from the MP2 component.

The corresponding results for $\delta_{MP2}^{CCSD(T)}$ in an aTZ basis set are presented in Figure 20. These results are only for the smallest eleven complexes in the S22 database.

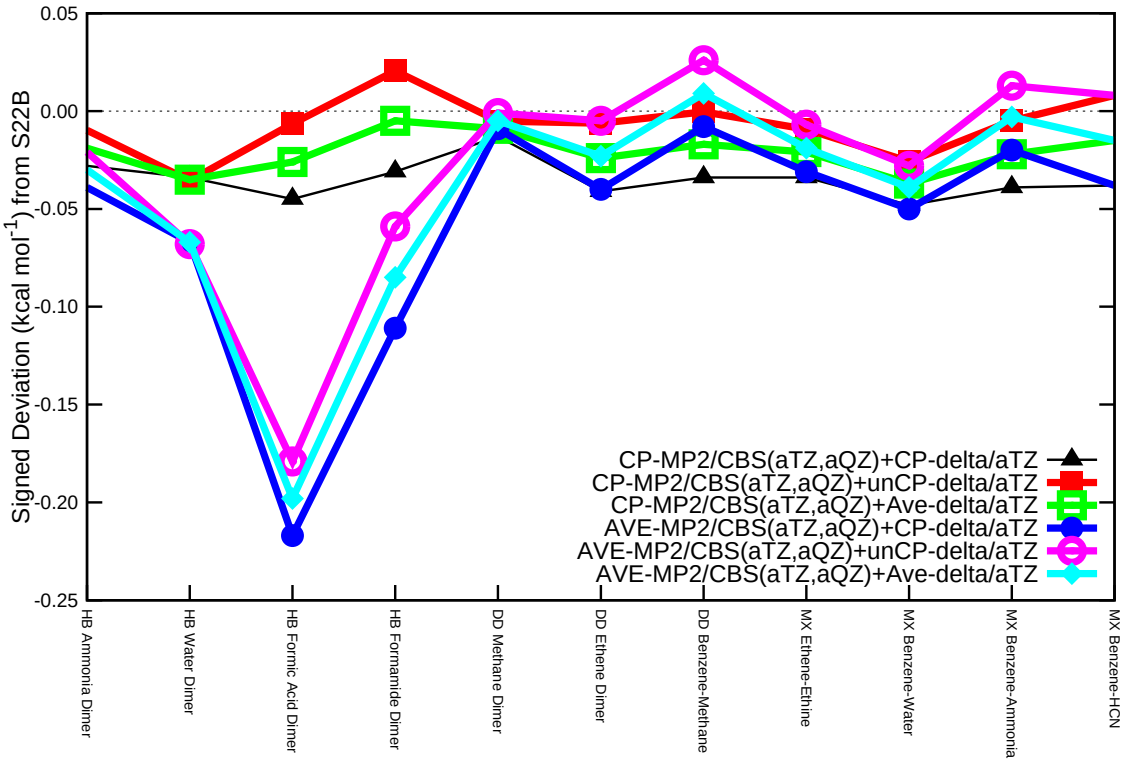

Figure 20: Signed errors of interaction energies versus the 11 smallest complexes in the S22B benchmark CCSD(T)/CBS values (Ref. [134]) for composite approaches using aug-cc-pVTZ.

This subset is used to ensure the reference data is significantly higher in quality than the methods being tested. CP-MP2/CBS(aTZ,aQZ) + $\delta_{CP-MP2}^{CP-CCSD(T)}$ achieves MAEs of (0.035, 0.030, 0.040, and 0.035 kcal mol^{-1}), while CP-MP2/CBS(aTZ,aQZ) + $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ achieves MAEs of (0.018, 0.004, 0.012, and 0.012 kcal mol^{-1}). This illustrates that $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ can provide favorable error cancellation in an aTZ basis set. CP-MP2/CBS(aTZ,aQZ) + $\delta_{Ave-MP2}^{Ave-CCSD(T)}$ achieves MAEs of (0.021, 0.017, 0.023, and 0.021 kcal mol^{-1}), which is in between the average performance of CP and unCP $\delta_{MP2}^{CCSD(T)}$ corrections. With CP-MP2/CBS(aTZ,aQZ), any of the three $\delta_{MP2}^{CCSD(T)}$ approaches seem to provide reliable benchmark quality results. When starting with an averaged MP2 reference instead of a CP-MP2 reference, significant error

is introduced, particularly in the hydrogen bound complexes. The largest error introduced is for formic acid dimer, where we report a deviation of 0.217 kcal mol^{-1} for Ave-MP2/CBS(aTZ,aQZ) with a $\delta_{CP-MP2}^{CP-CCSD(T)}$ corrections compared to an error of 0.045 for CP-MP2/CBS(aTZ,aQZ) with a $\delta_{CP-MP2}^{CP-CCSD(T)}$ correction. This introduced error comes from the remaining 0.441 kcal mol^{-1} BSSE in the MP2/CBS(aTZ,aQZ) computation, in which case the CP-MP2 outperforms both unCP-MP2 and Ave-MP2.

5.4 Conclusions

In conclusion, this work has shown that averaging CP and unCP corrected values does have merit for the MP2 reference when using aQZ quality basis sets, but averaging becomes less useful when using complete basis set extrapolations such as CBS(aQZ,a5Z) because the BSSE is often abated to around 1% or less of the interaction energy as shown in several small complexes. Averaging the $\delta_{MP2}^{CCSD(T)}$ values in an aDZ basis set is not recommended because for each binding motif, either unCP (hydrogen bound) or CP (dispersion) perform best, so any mixing would introduce error. When employing the aTZ basis set, CP-MP2/CBS(aTZ,aQZ) with an $\delta_{unCP-MP2}^{unCP-CCSD(T)}$ correction outperforms all other composite approaches overall, as well as for each binding motif individually. With this analysis, we have demonstrated that the approach proposed by Mackie and DiLabio[130] is not recommended as a black-box method for non-covalent interactions, and have provided error analysis for each composite approach against the S22 and S11 benchmark databases.

The ongoing debate of when to use CP corrections and when not remains to be concluded, but we have demonstrated here that it will depend highly on the binding motif. This illustrates the importance of using symmetry adapted perturbation theory (SAPT)[98] to understand the type of binding one is working with before proceeding. If at least aDZ quality basis sets cannot be afforded for the $\delta_{MP2}^{CCSD(T)}$ corrections, we do not advise using focal-point schemes as we have recently shown that this kind

of approach can often lead to large uncertainty in the estimate of the interaction energy.[134]

CHAPTER VI

DISPERSION-WEIGHTED EXPLICITLY CORRELATED COUPLED-CLUSTER THEORY: DW-CCSD(T**)-F12

6.1 Abstract

We propose a procedure denoted dispersion-weighted explicitly-correlated coupled-cluster [DW-CCSD(T**)-F12] which mixes CCSD(T**)-F12a and CCSD(T**)-F12b so as to correct the small errors exhibited by each of the approximations in a small basis set, allowing for a black-box method that can provide high-quality interaction energies for a variety of non-bonded interactions. Relative to CCSD(T**)-F12a and CCSD(T**)-F12b, DW-CCSD(T**)-F12 reduces the mean absolute deviation by a factor of 2 and the maximum error by a factor of 3 (formic acid dimer) and 4 (stacked adenine-thymine) for the aug-cc-pVDZ basis set.

6.2 Introduction

Coupled-cluster with single, double, and perturbative triple excitations [CCSD(T)] has been referred to as the "gold standard" in computational chemistry.[124] This sophisticated description of dynamic electron correlation is often required to achieve benchmark quality results. In the context of non-covalent interactions, CCSD(T) can provide very accurate results,[93] but this typically requires very large basis sets, augmented with diffuse functions. Coupled with the steep computational scaling of CCSD(T), this significantly restricts the size of systems that may be studied at this level of accuracy. With the introduction of explicitly correlated wavefunctions, however, this computational cost is severely abated because accurate energies may be attained using relatively small basis sets.[138, 113, 230, 131, 217, 229, 4, 231, 114, 116,

194, 40] Applying these methods to non-covalent interactions, Marchetti *et al.*[132] showed that CCSD(T**)-F12a/aug-cc-pVDZ can achieve an accuracy of better than 0.2 kcal mol^{-1} for all dimers in the S22 test set.[103] They argued that this same accuracy would require at least an aug-cc-pVQZ basis set with traditional CCSD(T), which would be 1-2 orders of magnitude more expensive. More recently, de Lange *et al.*[40] demonstrated that CCSD(T)-F12a/VDZ-F12 and CCSD(T)-F12b/VDZ-F12 achieve an average accuracy of 0.03 kcal mol^{-1} compared to CCSD(T)/CBS(a5Z,a6Z) for small molecules interacting with carbon dioxide. This accuracy is impressive considering the standard CCSD(T)/aug-cc-pVDZ has an average error of 0.4 kcal mol^{-1} for these complexes.

When employing explicitly correlated methods, the choice of ansatz and how to treat the triples correction are both very important. For coupled-cluster, the F12a and F12b approximations have become the most commonly used.[4, 116] Within the MOLPRO package[2] used here, since there is no explicitly correlated triples correction, this quantity should be scaled to achieve highly accurate results. This work follows the approach of Marchetti *et al.*[132] whereby the triples correction is scaled by the ratio of MP2 correlation energy and MP2-F12 correlation energy:

$$E^{corr}_{(T^*)} = E^{corr}_{(T)} \cdot \frac{E_{MP2-F12}}{E_{MP2}}. \tag{74}$$

To retain size-consistency for interaction energies, one must use the same scale factor for all computations (the dimer, and both monomers). Because of this difference, we designate CCSD(T**)-F12 (2 asterisks) to refer to the size-consistent version (using the dimer scale factor for all three computations), and CCSD(T*)-F12 (1 asterisk) to refer to the independently scaled version.

In our recent study,[134] CCSD(T**)-F12a/aug-cc-pVDZ is shown to yield a MAD of 0.09 kcal mol^{-1} against the newly revised interaction energies for the S22 test set[103] (herein referred to as S22B).[134] This accuracy is notable because standard

CCSD(T) (in the absence of focal-point analysis) would require a much larger basis set, such as aug-cc-pVQZ, to reach this accuracy, and such computations would become prohibitively expensive for all but the smallest molecular systems. Notably, CCSD(T**)-F12a reproduces benchmark energies for hydrogen bonded systems very accurately while incurring small errors for dispersion bound complexes, and CCSD(T**)-F12b reproduces benchmark energies for dispersion bound complexes very accurately but degrades somewhat in quality for hydrogen bonding. A very similar problem has been investigated by Marchetti *et al.*[132] in their studies of explicitly correlated spin-component-scaled Møller–Plesset perturbation theory (SCS-MP2-F12) and MP2-F12. In their work, they found MP2 significantly overbinds dispersion-bound complexes but does well for hydrogen bonding, while SCS-MP2-F12 can properly describe dispersion but underbinds hydrogen-bonding complexes. They proposed mixing the two approaches in a method known as dispersion-weighted MP2 (DW-MP2):

$$\Delta E_{DW-MP2} = \omega \Delta E_{MP2-F12} + (1-\omega)\Delta E_{SCS-MP2-F12}, \tag{75}$$

$$\omega = \frac{1}{2}[1 + tanh(a + b\frac{\Delta E_{SCF}}{\Delta E_{MP2-F12}})]. \tag{76}$$

The chosen switching function (equation 76) is a hyperbolic tangent function with two fit parameters, and the switching metric between SCS-MP2-F12 and a MP2-F12 is the ratio of MP2-F12 and self-consistent field (SCF) interaction energies. The underlying concept is that hydrogen bonded complexes have a ratio near one because the interaction is predominantly electrostatic, which is properly captured by SCF, while dispersion-dominated systems should yield a ratio far from one, as SCF fails to model dispersion because of its lack of dynamic electron correlation. This mixing transforms MP2-F12 [root-mean-square error (RMSE) of 1.17 kcal mol^{-1} for the aug-cc-pVDZ basis] and SCS-MP2-F12 [1.08 kcal mol^{-1}] into DW-MP2 [0.24 kcal mol^{-1}]

for the S22B test set. Inspired by this procedure we explore analogous methods of combining CCSD(T**)-F12a and CCSD(T**)-F12b based on the character of the noncovalent interaction.

6.3 Theoretical Methods

As demonstrated in Figures 21 and 22, direct computation of interaction energies by CCSD(T**)-F12, even when using the modest aug-cc-pVDZ basis set, yields surprisingly accurate values compared to the best available estimates.

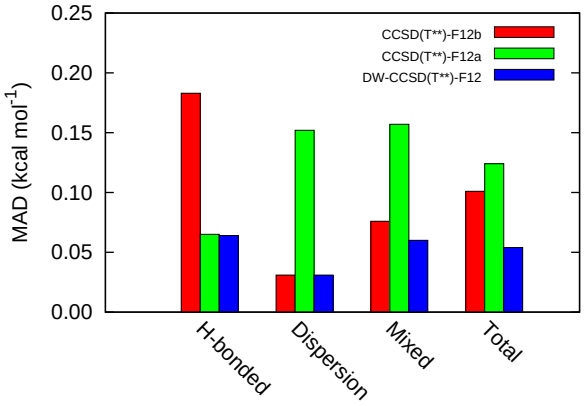

Figure 21: Mean absolute deviation (MAD) of interaction energies versus S22B benchmark CCSD(T)/CBS values (Ref. [134]) for explicitly correlated methods and DW-CCSD(T**)-F12 using the aug-cc-pVDZ basis.

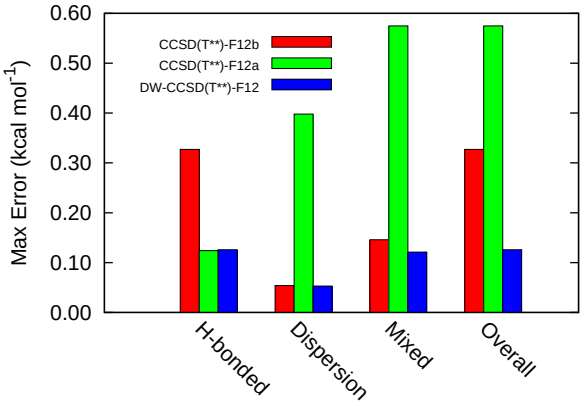

Figure 22: Maximum absolute deviation of interaction energies versus S22B benchmark CCSD(T)/CBS (Ref. [134]) values for explicitly correlated methods and DW-CCSD(T**)-F12 using the aug-cc-pVDZ basis.

Here, we seek to correct the minor remaining deficiencies in explicitly correlated CCSD(T**)-F12a and CCSD(T**)-F12b with a scheme similar to the DW-MP2 approach of Marchetti *et al.*[132]

$$\Delta E_{DW-CCSD(T^{**})-F12} = \omega \Delta E_{CCSD(T^{**})-F12a} + (1-\omega)\Delta E_{CCSD(T^{**})-F12b} \qquad (77)$$

$$\omega = \frac{1}{2}[1 + tanh(\alpha + \beta \frac{(\Delta E_{SCF} - \Delta E_{MP2-F12})}{\Delta E_{MP2-F12}})] \qquad (78)$$

Here, the SCF energy includes the complementary auxiliary orbital basis (CABS) correction.[224] Using Eqns 77 and 78, we fit α and β against the S22B test set. To test the transferability of these fit parameters, we also consider another test set, HSG-A.[134, 50]

The two test sets considered here, S22 and HSG, were chosen for their small size, convenient separation into hydrogen-bonded, dispersion-dominated, and mixed bonding classes, and quality of available benchmark interaction energies. S22 is an

established test set with systems ranging from water dimer to adenine·thymine complexes. Its reference values have recently been revised (S22B) to be of MP2/CBS(aug-cc-pVTZ,aug-cc-pVQZ) + $\delta_{MP2}^{CCSD(T)}$/aug-cc-pVTZ quality or better. The HSG test set was formed by dissecting the binding site of a bound protein-drug complex (HIV-II protease-indinavir) into twenty-one pairs of chemical fragments (each of which are not necessarily at its individual equilibrium geometries).[50] We recently revised[134] the HSG benchmark energies replacing the previous extrapolated heavy-aug-cc-pVDZ/heavy-aug-cc-pVTZ $\delta_{MP2}^{CCSD(T)}$ corrections with the bare heavy-aug-cc-pVTZ $\delta_{MP2}^{CCSD(T)}$ corrections to avoid overestimating the coupled-cluster correction for hydrogen-bonded systems [here heavy-aug-cc-pVXZ refers to the aug-cc-pVXZ basis where diffuse functions are added only to heavy (non-hydrogen) atoms]. This revision results from the recent systematic study of the non-monotonic convergence of $\delta_{MP2}^{CCSD(T)}$ for various members of the S22 test set.[134] The revised HSG-A values are thus similar accuracy to the S22B values.

6.4 Results and Discussion

Using equations 77 and 78, we smoothly combine the F12a to F12b ansatze according to the relative difference in ΔE_{SCF} and $\Delta E_{MP2-F12}$ interaction energies. Figure 21 and Table 11 illustrate the errors associated with each ansatz independently, and it is clear that some combination of the two should result in a method that captures both types of bonding accurately on average.

Table 11: Counterpoise corrected interaction energies (kcal mol^{-1}) for the CCSD(T**)-F12a, CCSD(T**)-F12b, and DW-CCSD(T**)-F12 methods compared to S22B benchmark values (Ref. [134]). Fit parameters are α=3 and β=4. Computations use the aug-cc-pVDZ basis set. The errors and mixing in the DW approach can be seen graphically in Figure 23.

	Complex	**Reference IE**	**F12a**	**F12b**	**DW-F12**
1	HB Ammonia Dimer, C_{2h}	-3.133	-3.11	-3.05	-3.10
2	HB Water Dimer, C_s	-4.989	-4.92	-4.86	-4.92
3	HB Formic Acid Dimer, C_{2h}	-18.753	-18.63	-18.43	-18.63
4	HB Formamide Dimer, C_{2h}	-16.062	-15.94	-15.80	-15.94
5	HB Hydrogen-Bonded Uracil Dimer, C_{2h}	-20.641	-20.63	-20.45	-20.63
6	HB 2-Pyridone·2-Aminopyridine, C_1	-16.934	-16.98	-16.79	-16.97
7	HB Adenine·Thymine WC, C_1	-16.660	-16.72	-16.51	-16.71
8	DD Methane Dimer, D_{3d}	-0.527	-0.53	-0.51	-0.51
9	DD Ethene Dimer, D_{2d}	-1.472	-1.50	-1.44	-1.44
10	DD Benzene·Methane, C_3	-1.448	-1.47	-1.40	-1.40
11	DD Parallel Displaced Benzene Dimer, C_{2h}	-2.654	-2.90	-2.64	-2.64
12	DD Pyrazine Dimer, C_s	-4.255	-4.54	-4.26	-4.26
13	MX Stacked Uracil Dimer, C_2	-9.805	-10.17	-9.84	-9.88
14	DD Stacked Indole·Benzene, C_1	-4.524	-4.92	-4.57	-4.57
15	MX Stacked Adenine·Thymine, C_1	-11.730	-12.30	-11.84	-11.85
16	MX Ethene·Ethine, C_{2v}	-1.496	-1.51	-1.47	-1.49
17	MX Benzene·Water, C_s	-3.275	-3.23	-3.13	-3.19
18	MX Benzene·Ammonia, C_s	-2.312	-2.31	-2.23	-2.23
19	MX Benzene·Hydrogen Cyanide, C_s	-4.541	-4.49	-4.39	-4.47
20	DD T-shaped Benzene Dimer, C_{2v}	-2.717	-2.78	-2.66	-2.66
21	MX T-shaped Indole·Benzene, C_1	-5.627	-5.73	-5.56	-5.58
22	MX Phenol Dimer, C_1	-7.097	-7.15	-6.99	-7.07
	Hydrogen Bonded				
	Maximal Deviation		0.12	0.33	0.13
	Mean Signed Deviation		0.03	0.18	0.04
	Mean Absolute Deviation		0.06	0.18	0.06
	Root-Mean-Square Deviation		0.08	0.20	0.08
	Mixed Influence				
	Maximal Deviation		-0.57	0.15	-0.12
	Mean Signed Deviation		-0.13	0.05	0.02
	Mean Absolute Deviation		0.15	0.09	0.07
	Root-Mean-Square Deviation		0.25	0.10	0.07
	Dispersion Bound				
	Maximal Deviation		-0.01	0.05	0.05
	Mean Signed Deviation		-0.40	0.02	0.02
	Mean Absolute Deviation		0.15	0.03	0.03
	Root-Mean-Square Deviation		0.21	0.04	0.04
	Full Set				
	Maximal Deviation		-0.57	0.33	0.13
	Mean Signed Deviation		-0.08	0.08	0.02
	Mean Absolute Deviation		0.12	0.10	0.05
	Root-Mean-Square Deviation		0.19	0.13	0.07

We find fitted parameters of α=3 and β=4 for the S22B set, which results in a MAD of 0.06, 0.03, 0.07, and 0.05 kcal mol^{-1}, respectively, for hydrogen-bonding, dispersion, mixed bonding, and overall. The complexes were assigned to these categories according to the SAPT2+(3)/aug-cc-pVTZ results of Hohenstein and Sherrill.[85] The most substantial gain of this approach is in the reduction of maximum errors for each subset, as shown in Figure 22. The maximum errors for F12a and F12b are 0.58 and 0.33 kcal mol^{-1}, respectively, whereas the largest error incurred by DW-CCSD(T**)-F12 is 0.13 kcal mol^{-1} for formamide dimer. This is a factor of 3-4 reduction in the maximum error and a factor of 2 reduction in the MAD. Figure 23 shows the error for each complex in the S22B test set for CCSD(T**)-F12a, CCSD(T**)-F12b, and DW-CCSD(T**)-F12.

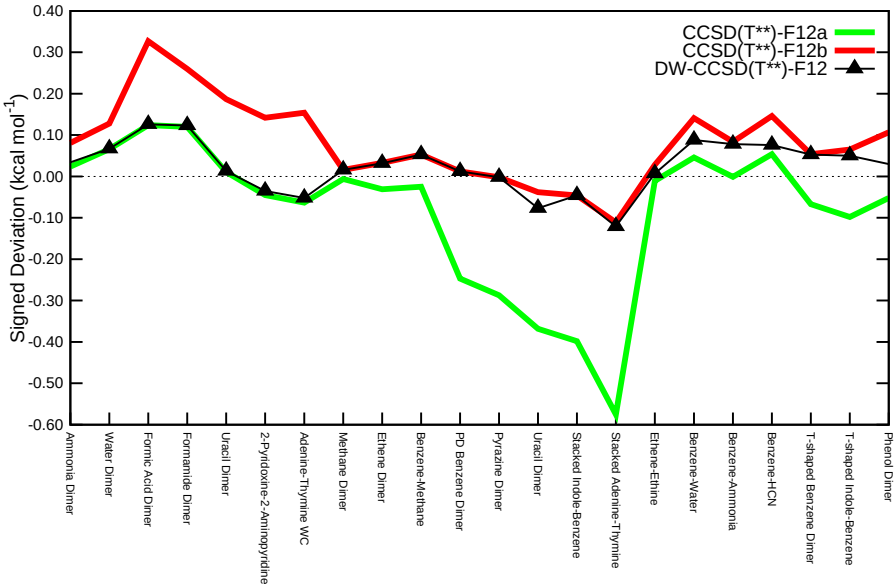

Figure 23: For the aug-cc-pVDZ basis set, interaction energy errors for CCSD(T**)-F12a, CCSD(T**)-F12b, and DW-CCSD(T**)-F12 methods for each complex in the S22 test set. All errors in kcal mol^{-1}, relative to S22B CCSD(T)/CBS benchmarks (Ref. [134]). Individual errors and statistics can be found in Table 11.

In this figure, one can clearly see how DW-CCSD(T**)-F12 switches between CCSD(T**)-F12a and CCSD(T**)-F12b to avoid the large maximum errors.

To evaluate the transferability of fitting parameters, DW-CCSD(T**)-F12/aug-cc-pVDZ was applied to the HSG test set using the parameters fit against the S22B test set; it achieves a MAD of 0.04 kcal mol^{-1} for overall interaction energies. This is a significant achievement, especially because the diverse, non-equilibrium nature of the complexes found in the HSG set is often challenging for computational methods. The maximum error across this test set is 0.10 kcal mol^{-1}. These impressive results compare to a MAD of 0.42 kcal mol^{-1} and a maximum error of 1.53 kcal mol^{-1} for conventional CCSD(T)/heavy-aug-cc-pVDZ.

DW-CCSD(T**)-F12 was also tested using an aug-cc-pVTZ basis set against the ten smallest complexes in the S22B benchmark set (Figure 24 and Table 12).

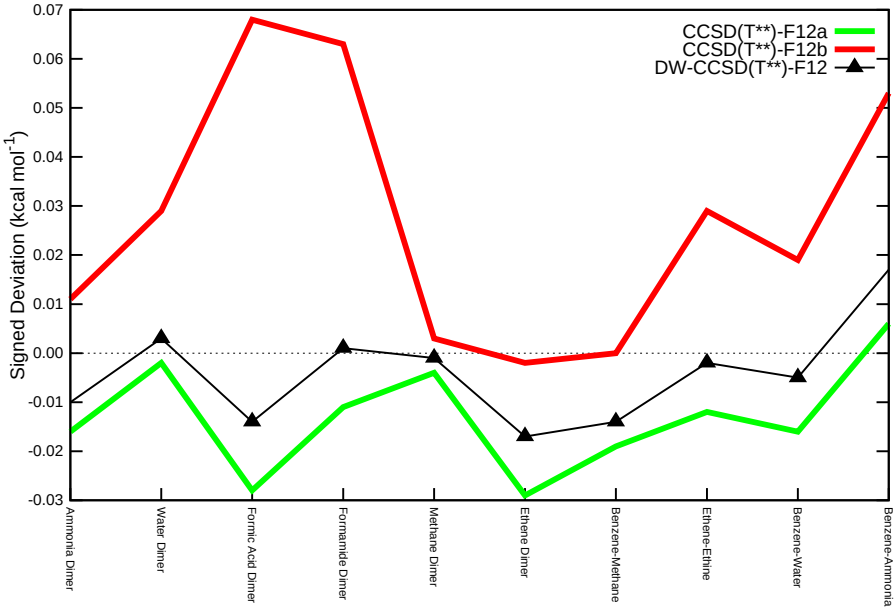

Figure 24: For the aug-cc-pVTZ basis set, interaction energy errors for CCSD(T**)-F12a, CCSD(T**)-F12b, and DW-CCSD(T**)-F12 methods for each complex in the S22 test set. All errors in kcal mol^{-1}, relative to a subset of the S22B CCSD(T)/CBS benchmarks (Ref. [134]).

Table 12: Counterpoise corrected interaction energies (kcal mol^{-1}) for the CCSD(T**)-F12a, CCSD(T**)-F12b, and DW-CCSD(T**)-F12 methods compared to S22B benchmark values (Ref. [134]). Fit parameters are α=1 and β=0.6. Computations use the aug-cc-pVTZ basis set. The errors and mixing in the DW approach can be seen graphically in Figure 24.

	Complex	**Reference IE**	**F12a**	**F12b**	**DW-F12**
1	HB Ammonia Dimer, C_{2h}	-3.133	-3.15	-3.12	-3.14
2	HB Water Dimer, C_s	-4.989	-4.99	-4.96	-4.99
3	HB Formic Acid Dimer, C_{2h}	-18.753	-18.78	-18.69	-18.77
4	HB Formamide Dimer, C_{2h}	-16.062	-16.07	-16.00	-16.06
8	DD Methane Dimer, D_{3d}	-0.527	-0.53	-0.52	-0.53
9	DD Ethene Dimer, D_{2d}	-1.472	-1.50	-1.47	-1.49
16	MX Ethene·Ethine, C_{2v}	-1.496	-1.52	-1.50	-1.51
17	MX Benzene·Water, C_s	-3.275	-3.29	-3.25	-3.28
18	MX Benzene·Ammonia, C_s	-2.312	-2.33	-2.29	-2.32
19	MX Benzene·Hydrogen Cyanide, C_s	-4.541	-4.54	-4.49	-4.53
	Maximal Deviation		0.068	0.029	0.017
	Mean Absolute Deviation		0.028	0.014	0.008

With fit parameters α=1 and β=0.6, we find MADs of 0.028, 0.014, and 0.008 kcal mol^{-1} for CCSD(T**)-F12a, CCSD(T**)-F12b, and DW-CCSD(T**)-F12 respectively. These methods achieve maximum errors of 0.068, 0.029, and 0.017 kcal mol^{-1} for CCSD(T**)-F12a, CCSD(T**)-F12b, and DW-CCSD(T**)-F12 respectively. When using the aug-cc-pVDZ fit parameters of α=3 and β=4, DW-CCSD(T**)-F12 still achieves an overall MAD of 0.011 kcal mol^{-1} and a maximum error of 0.027 kcal mol^{-1} (Formic acid dimer).

6.5 Conclusions

A new "dispersion-weighted" approach is proposed for explicitly correlated coupled-cluster studies of weakly bound systems that accurately describes both hydrogen-bonding and dispersion bound complexes through a black-box admixture of CCSD(T**)-F12a and CCSD(T**)-F12b. This technique achieves MAD values of 0.05 and 0.04 kcal mol^{-1} for the S22B and HSG-A test sets (using S22B-fit parameters), corresponding to a factor of 2 reduction of the MAD and a factor of 3 and 4 reduction in maximum error relative to F12a and F12b methods, respectively. Explicitly correlated CCSD-F12 computations are more expensive than canonical CCSD, but the extra cost is usually negligible with the inclusion of perturbative triples. This suggests that CCSD(T**)-F12, and perhaps the dispersion-weighted variant proposed here, should become a preferred approach for obtaining accurate benchmarks for non-covalent interactions.

CHAPTER VII

CONCLUSIONS AND OUTLOOK

This work demonstrates how one would define "benchmark" quality computations for non-bonded interactions and how benchmark quality databases can stimulate new understanding of non-covalent interactions. We have shown through high-level computations that certain interactions, such as cation-π interactions, cannot be modeled well by common force-fields because of their lack of a proper description of induction. The class of π-π interactions has also been shown to be troublesome for common force-fields because of their use of atom-centered charges, which fail to properly describe complexes where π-clouds are overlapping. While benchmark quality computations such as CCSD(T) are often very reliable black-box methods, this procedure can yet have large error bounds when not used properly. We have shown that for double-ζ basis sets commonly used in the literature, errors on the order of 0.10–0.39 kcal mol^{-1} can occur for small van der Waals dimers. We also illustrated that complete basis set (CBS) extrapolations of the $\delta_{MP2}^{CCSD(T)}$ coupled-cluster correction using the aug-cc-pVDZ and aug-cc-pVTZ basis set can introduce significant amounts of error, upwards of 0.26 kcal mol^{-1} for the formic acid dimer in the HBC6 database. This result was surprising and still requires more detailed examination to better explain why CBS extrapolations perform so poorly even with a generous CBS(aTZ,aQZ) basis set extrapolation. This work should prevent future papers from claiming "benchmark quality" results when their methods can clearly be shown to have significant errors from the statistics I have provided. This thesis also provides error statistics for each contribution to focal-point methods individually, as well as providing a detailed analysis of common approximations to those contributions. For the MP2 component,

we have shown an array of approximations from density-fitting to the use of heavy-augmented functions and how these approximations affect the overall accuracy as well as the magnitude of computational savings. These statistics allow researchers to continue to apply approximations until the desired speed and accuracy balance has been reached, resulting in fast, reliable benchmarks. We showed that with these very accurate databases, new methods can be fit to this data directly, improving accuracy empirically. In particular, this was demonstrated with the development of the DW-CCSD(T**)-F12 approach which mixes CCSD(T**)-F12a and CCSD(T**)-F12b energies so as to minimize the mean and maximum error of the interaction energies against the S22 dataset. We verified that the resulting fit parameters are transferable by applying this new method to other databases, such as HSG. With these simple fitting procedures, we have shown that the mean error can be reduced by a factor of two and the maximum error by a factor of three when this method is used in conjunction with the aug-cc-pVDZ basis set.

Because the accuracy of our "gold standard" benchmarks is so high, we are no longer constrained to only looking at deficiencies in low-level force-fields and semi-empirical methods but can now also examine the performance of wavefunction-based methods. In an upcoming paper with Dr. Lori Burns, we provide detailed statistics on how well each wavefunction-based method does with a variety of basis sets compared to our CCSD(T) benchmarks for 345 complexes (union of S22, NBC, HBC, and HSG datasets). With this knowledge, we can provide the community with a detailed performance versus accuracy analysis. As a result of this work, we have defined a hierarchy of methods that balance accuracy and computational cost. We define these new methods: *bronze*, which is MP2C-F12/aDZ; *silver*, which is DW-CCSD(T**)-F12/aDZ; and *gold*, which is estimated CCSD(T)/CBS. This hierarchical treatment of benchmarks allows us to continually improve our reference data with time.

With the hierarchy of *bronze*, *silver*, and *gold*, we can now apply these benchmark-quality methods to much larger scale problems. One such problem that we are working on at the moment is that of protein folding. In collaboration with Dr. Kenneth Merz's group from the University of Florida, we have built a database that includes many thousands of interactions important to protein folding, broken down into three groups: sidechain-sidechain, backbone-backbone, and sidechain-backbone. Using crystal structures from the PDB (with some minimum structural accuracy standards), we have constructed a database by fragmenting proteins into about ten thousand important interactions. We have started applying our newly developed hierarchy of benchmarks; *bronze* is complete for the majority and *silver* is in progress. Even at this early stage, we can provide very accurate statistics on the performance of various methods of interest to the modeling community. Of particular interest to us are examples where methods are performing very poorly (absolute errors of +15-20 kcal mol^{-1} in some cases), because it is through these examples that we learn why various low scaling methods are failing. We can take these select systems and pinpoint what physics is missing and hopefully provide insight into how to correct for it cheaply. The other benefit to these very large databases is for parameterizing newly developed methods. With the vast array of binding types at estimated CCSD(T)/CBS, any method that can fit to these interaction energies properly should be able to perform very well for protein folding problems in general.

One of the remaining problems to be elucidated in the area of non-bonded interactions is the inclusion of higher order excitation corrections. We have shown that $\delta_{MP2}^{CCSD(T)}$ corrections are very important, but no one has fully investigated the importance of including $\delta_{CCSD(T)}^{CCSDT(Q)}$ corrections or beyond. Hopkins *et al.*[93] estimated $\delta_{CCSD(T)}^{CCSD(TQ)}$ corrections to be approximately a tenth the magnitude of $\delta_{MP2}^{CCSD(T)}$ (with the same sign), but these were for relatively small systems. None of the few studies to date have convincingly demonstrated the importance of higher order corrections,

or the lack thereof. Another largely unexplored area of non-covalent interactions is solvation effects. Studying the effects of solvation on non-bonded interactions is an arduous task because of the sampling required, the number of solvent molecules that need to be included, and the lack of code with proper/fast derivatives to carry out such a computation. In a soon to be released letter, we show that implicit solvation models do not perform well enough to answer these very important questions. It should also be noted that this thesis has focused primarily on purely electronic interaction energies, but to compare to experiments properly requires computing ΔH and ΔS, which will launch this whole benchmarking to build upon this thesis. Over time we will be able to estimate each important contribution to compare to experiment with well defined error bars.

APPENDIX A

ANCILLARY MATERIAL

Table 13: Absolute error and absolute percent errors (in parenthesis) relative to CCSD(T)/ CBS interaction energies from the S22B test set for various small basis sets commonly used in computing $\delta_{MP2}^{CCSD(T)}$ corrections. Relative errors evaluated as a percentage of the overall CCSD(T)/ CBS interaction energies. All values are in kcal mol^{-1}.

Complex	S22B $\delta_{MP2}^{CCSD(T)}$	6-31G*		6-31G*(0.25)		6-31G**(0.25,0.1)		cc-pVDZ		aug-cc-pVDZ	
1	0.013	0.159	(5.1)	0.030	(0.9)	0.021	(0.7)	0.130	(4.1)	0.044	(1.4)
2	-0.039	0.408	(8.2)	0.109	(2.2)	0.034	(0.7)	0.276	(5.5)	0.073	(1.5)
3	-0.250	0.981	(5.2)	0.047	(0.3)	0.011	(0.1)	0.475	(2.5)	0.289	(1.5)
4	-0.290	0.603	(3.8)	0.043	(0.3)	0.084	(0.5)	0.341	(2.1)	0.225	(1.4)
5	-0.271	0.602	(2.9)	0.136	(0.7)	0.190	(0.9)	0.304	(1.5)	0.208	(1.0)
6	0.406	0.613	(3.6)	0.175	(1.0)	0.103	(0.6)	0.330	(1.9)	0.224	(1.3)
7	-0.140	0.575	(3.5)	0.100	(0.6)	0.021	(0.1)	0.308	(1.8)	0.217	(1.3)
8	-0.036	0.020	(3.7)	0.042	(7.9)	0.007	(1.4)	0.039	(7.3)	0.001	(0.1)
9	0.101	0.002	(0.1)	0.047	(3.2)	0.039	(2.6)	0.046	(3.1)	0.002	(0.1)
10	0.364	0.146	(10.1)	0.029	(2.0)	0.071	(4.9)	0.037	(2.6)	0.016	(1.1)
11	2.301	0.873	(33.0)	0.320	(12.1)	0.319	(12.0)	0.488	(18.4)	0.077	(2.9)
12	2.658	0.695	(16.4)	0.376	(8.9)	0.411	(9.7)	0.441	(10.4)	0.091	(2.1)
13	1.305	0.178	(1.8)	0.433	(4.4)	0.526	(5.4)	0.148	(1.5)	0.043	(0.4)
14	3.557	1.184	(26.2)	0.459	(10.1)	0.464	(10.3)	0.660	(14.6)	0.098	(2.2)
15	3.130	0.660	(5.6)	0.630	(5.4)	0.754	(6.4)	0.431	(3.7)	0.451	(3.9)
16	0.161	0.026	(1.7)	0.045	(3.0)	0.008	(0.5)	0.046	(3.0)	0.003	(0.2)
17	0.262	0.065	(2.0)	0.001	(0.0)	0.093	(2.8)	0.081	(2.5)	0.005	(0.2)
18	0.334	0.063	(2.7)	0.007	(0.3)	0.073	(3.1)	0.012	(0.5)	0.004	(0.2)
19	0.628	0.034	(0.7)	0.003	(0.1)	0.054	(1.2)	0.057	(1.3)	0.010	(0.2)
20	0.909	0.271	(10.0)	0.092	(3.4)	0.129	(4.7)	0.101	(3.7)	0.028	(1.0)
21	1.348	0.246	(4.4)	0.062	(1.1)	0.150	(2.7)	0.051	(0.9)	0.004	(0.1)
22	0.643	0.130	(1.8)	0.051	(0.7)	0.126	(1.8)	0.084	(1.2)	0.051	(0.7)
MAD	-	0.388	(6.9)	0.147	(3.1)	0.168	(3.3)	0.222	(4.3)	0.098	(1.1)

Table 14: NBC10A benchmark interaction energies [using HF/aQZ, MP2/ CBS(aTZ,aQZ) extrapolation, and $\delta_{MP2}^{CCSD(T)}$/haTZ components] and differences (kcal mol^{-1}) with respect to original values [using $\delta_{MP2}^{CCSD(T)}$/ CBS(haDZ,haTZ)] for the sandwich (S) and T-shaped (T) benzene dimers at various intermolecular distances (in Å).

Complex	**Original**[a]	**NBC10A**	| Δ |
BzBz_S-3.2	3.522	3.462	0.060
BzBz_S-3.3	1.535	1.484	0.051
BzBz_S-3.4	0.189	0.147	0.042
BzBz_S-3.5	-0.689	-0.724	0.035
BzBz_S-3.6	-1.231	-1.259	0.028
BzBz_S-3.7	-1.535	-1.558	0.023
BzBz_S-3.8	-1.674	-1.693	0.019
BzBz_S-3.9	-1.701	-1.717	0.016
BzBz_S-4.0	-1.655	-1.669	0.014
BzBz_S-4.1	-1.565	-1.577	0.012
BzBz_S-4.2	-1.448	-1.459	0.011
BzBz_S-4.5	-1.058	-1.066	0.008
BzBz_S-5.0	-0.542	-0.546	0.004
BzBz_S-5.5	-0.248	-0.251	0.003
BzBz_S-6.0	-0.099	-0.101	0.002
BzBz_S-6.5	-0.028	-0.029	0.001
BzBz_S-10.0	0.018	0.018	0.000
BzBz_T-4.4	0.626	0.617	0.009
BzBz_T-4.5	-0.760	-0.769	0.009
BzBz_T-4.6	-1.673	-1.682	0.009
BzBz_T-4.7	-2.239	-2.246	0.007
BzBz_T-4.8	-2.552	-2.559	0.007
BzBz_T-4.9	-2.687	-2.693	0.006
BzBz_T-5.0	-2.698	-2.703	0.005
BzBz_T-5.1	-2.627	-2.630	0.003
BzBz_T-5.2	-2.503	-2.506	0.003
BzBz_T-5.3	-2.349	-2.351	0.002
BzBz_T-5.4	-2.179	-2.181	0.002
BzBz_T-5.5	-2.005	-2.006	0.001
BzBz_T-5.6	-1.833	-1.834	0.001
BzBz_T-6.0	-1.242	-1.242	0.000
BzBz_T-6.5	-0.752	-0.752	0.000
BzBz_T-7.0	-0.468	-0.468	0.000
BzBz_T-7.5	-0.302	-0.302	0.000
BzBz_T-8.0	-0.203	-0.203	0.000

Table 15: NBC10A benchmark interaction energies [using HF/aQZ, MP2/ CBS(aTZ,aQZ) extrapolation, and $\delta_{MP2}^{CCSD(T)}$/haTZ components] and differences (kcal mol^{-1}) with respect to original values [using $\delta_{MP2}^{CCSD(T)}$/ CBS(haDZ,haTZ)] for the parallel-displaced (PD) benzene dimer at various horizontal displacements for fixed vertical displacement of 3.2 Å.

Complex	**Original**[a]	**NBC10A**	| Δ |
BzBz_PD32-0.2	3.301	3.241	0.060
BzBz_PD32-0.4	2.678	2.619	0.059
BzBz_PD32-0.6	1.783	1.726	0.057
BzBz_PD32-0.8	0.781	0.726	0.055
BzBz_PD32-1.0	-0.171	-0.222	0.051
BzBz_PD32-1.2	-0.954	-1.002	0.048
BzBz_PD32-1.4	-1.508	-1.553	0.045
BzBz_PD32-1.5	-1.695	-1.738	0.043
BzBz_PD32-1.6	-1.827	-1.868	0.041
BzBz_PD32-1.7	-1.911	-1.949	0.038
BzBz_PD32-1.8	-1.950	-1.988	0.038
BzBz_PD32-1.9	-1.957	-1.992	0.035
BzBz_PD32-2.0	-1.937	-1.971	0.034
BzBz_PD32-2.2	-1.860	-1.891	0.031
BzBz_PD32-2.4	-1.767	-1.795	0.028
BzBz_PD32-2.6	-1.702	-1.727	0.025
BzBz_PD32-2.8	-1.680	-1.702	0.022
BzBz_PD32-3.0	-1.705	-1.725	0.020

Table 16: NBC10A benchmark interaction energies [using HF/aQZ, MP2/ CBS(aTZ,aQZ) extrapolation, and $\delta_{MP2}^{CCSD(T)}$/haTZ components] and differences (kcal mol^{-1}) with respect to original values [using $\delta_{MP2}^{CCSD(T)}$/ CBS(haDZ,haTZ)] for the parallel-displaced (PD) benzene dimer at various horizontal displacements for fixed vertical displacement of 3.4 Å.

Complex	**Original**[a]	**NBC10A**	| Δ |
BzBz_PD34-0.2	0.070	0.029	0.041
BzBz_PD34-0.4	-0.257	-0.298	0.041
BzBz_PD34-0.6	-0.728	-0.768	0.040
BzBz_PD34-0.8	-1.260	-1.298	0.038
BzBz_PD34-1.0	-1.766	-1.802	0.036
BzBz_PD34-1.2	-2.179	-2.213	0.034
BzBz_PD34-1.4	-2.466	-2.497	0.031
BzBz_PD34-1.5	-2.557	-2.586	0.029
BzBz_PD34-1.6	-2.614	-2.643	0.029
BzBz_PD34-1.7	-2.640	-2.668	0.028
BzBz_PD34-1.8	-2.643	-2.670	0.027
BzBz_PD34-1.9	-2.624	-2.649	0.025
BzBz_PD34-2.0	-2.587	-2.611	0.024
BzBz_PD34-2.2	-2.479	-2.501	0.022
BzBz_PD34-2.4	-2.356	-2.377	0.021
BzBz_PD34-2.6	-2.242	-2.260	0.018
BzBz_PD34-2.8	-2.147	-2.163	0.016
BzBz_PD34-3.0	-2.079	-2.093	0.014

Table 17: NBC10A benchmark interaction energies [using HF/aQZ, MP2/ CBS(aTZ,aQZ) extrapolation, and $\delta_{MP2}^{CCSD(T)}$/haTZ components] and differences (kcal mol^{-1}) with respect to original values [using $\delta_{MP2}^{CCSD(T)}$/ CBS(haDZ,haTZ)] for the parallel-displaced (PD) benzene dimer at various horizontal displacements for fixed vertical displacement of 3.6 Å.

Complex	**Original**[a]	**NBC10A**	| Δ |
BzBz_PD36-0.2	-1.293	-1.321	0.028
BzBz_PD36-0.4	-1.462	-1.490	0.028
BzBz_PD36-0.6	-1.708	-1.735	0.027
BzBz_PD36-0.8	-1.984	-2.011	0.027
BzBz_PD36-1.0	-2.248	-2.273	0.025
BzBz_PD36-1.2	-2.458	-2.482	0.024
BzBz_PD36-1.4	-2.597	-2.619	0.022
BzBz_PD36-1.5	-2.635	-2.657	0.022
BzBz_PD36-1.6	-2.652	-2.674	0.022
BzBz_PD36-1.7	-2.654	-2.675	0.021
BzBz_PD36-1.8	-2.642	-2.662	0.020
BzBz_PD36-1.9	-2.615	-2.633	0.018
BzBz_PD36-2.0	-2.575	-2.593	0.018
BzBz_PD36-2.2	-2.473	-2.489	0.016
BzBz_PD36-2.4	-2.356	-2.371	0.015
BzBz_PD36-2.6	-2.240	-2.253	0.013
BzBz_PD36-2.8	-2.130	-2.143	0.013
BzBz_PD36-3.0	-2.035	-2.046	0.011

Table 18: NBC10A benchmark interaction energies [using HF/aQZ, MP2/ CBS(aTZ,aQZ) extrapolation, and $\delta_{MP2}^{CCSD(T)}$/aTZ components] and differences (kcal mol^{-1}) with respect to original values [using $\delta_{MP2}^{CCSD(T)}$/ CBS(aDZ,aTZ)] for the benzene · H_2S and benzene · methane complexes at various intermolecular distances (in Å).

Complex	**Original**[a]	**NBC10A**	| Δ |
BzH2S-3.2	1.250	1.236	0.014
BzH2S-3.4	-1.570	-1.584	0.014
BzH2S-3.5	-2.256	-2.269	0.013
BzH2S-3.6	-2.638	-2.649	0.011
BzH2S-3.7	-2.808	-2.818	0.010
BzH2S-3.8	-2.834	-2.843	0.009
BzH2S-3.9	-2.766	-2.773	0.007
BzH2S-4.0	-2.639	-2.645	0.006
BzH2S-4.1	-2.478	-2.483	0.005
BzH2S-4.2	-2.301	-2.305	0.004
BzH2S-4.5	-1.770	-1.771	0.001
BzH2S-4.75	-1.393	-1.393	0.000
BzH2S-5.0	-1.093	-1.092	0.001
BzH2S-5.25	-0.861	-0.859	0.002
BzH2S-5.5	-0.684	-0.682	0.002
BzH2S-6.0	-0.446	-0.444	0.002
BzH2S-6.5	-0.302	-0.301	0.001
BzH2S-7.0	-0.214	-0.212	0.002
BzH2S-7.5	-0.155	-0.154	0.001
BzMe-3.2	0.717	0.686	0.031
BzMe-3.3	-0.183	-0.213	0.030
BzMe-3.4	-0.774	-0.805	0.031
BzMe-3.5	-1.135	-1.173	0.038
BzMe-3.6	-1.337	-1.378	0.041
BzMe-3.7	-1.432	-1.470	0.038
BzMe-3.8	-1.439	-1.484	0.045
BzMe-3.9	-1.414	-1.445	0.031
BzMe-4.0	-1.327	-1.374	0.047
BzMe-4.1	-1.232	-1.284	0.052
BzMe-4.2	-1.138	-1.185	0.047
BzMe-4.4	-0.950	-0.984	0.034
BzMe-4.6	-0.760	-0.800	0.040
BzMe-4.8	-0.606	-0.643	0.037
BzMe-5.0	-0.475	-0.515	0.040
BzMe-5.2	-0.370	-0.413	0.043
BzMe-5.4	-0.286	-0.332	0.046
BzMe-5.6	-0.230	-0.268	0.038
BzMe-6.0	-0.141	-0.177	0.036

Table 19: NBC10A benchmark interaction energies [using HF/aQZ reference energies and directly extrapolated CCSD(T)/ CBS(aTZ,aQZ) correlation energies][a] (kcal mol^{-1}) for the methane dimer at various intermolecular distances (in Å). In this case, the original values are not changed in the NBC10A revision; they are repeated here for the convenience of the reader.

Complex	**Original**[a]	**NBC10A**	$\mid \Delta \mid$
MeMe-3.2	0.069	0.069	0.000
MeMe-3.3	-0.239	-0.239	0.000
MeMe-3.4	-0.417	-0.417	0.000
MeMe-3.5	-0.508	-0.508	0.000
MeMe-3.6	-0.541	-0.541	0.000
MeMe-3.7	-0.539	-0.539	0.000
MeMe-3.8	-0.515	-0.515	0.000
MeMe-3.9	-0.480	-0.480	0.000
MeMe-4.0	-0.439	-0.439	0.000
MeMe-4.1	-0.396	-0.396	0.000
MeMe-4.2	-0.354	-0.354	0.000
MeMe-4.3	-0.315	-0.315	0.000
MeMe-4.4	-0.279	-0.279	0.000
MeMe-4.6	-0.217	-0.217	0.000
MeMe-4.8	-0.168	-0.168	0.000
MeMe-5.0	-0.130	-0.130	0.000
MeMe-5.4	-0.080	-0.080	0.000
MeMe-5.8	-0.050	-0.050	0.000

[a]T. Takatani and C. D. Sherrill, *Phys. Chem. Chem. Phys.* **9**, 6106 (2007).

Table 20: NBC10A benchmark interaction energies [using HF/aQZ, MP2/ CBS(aTZ,aQZ) extrapolation, and $\delta_{MP2}^{CCSD(T)}$/haTZ components] and differences (kcal mol^{-1}) with respect to previous values [using $\delta_{MP2}^{CCSD(T)}$/ CBS(haDZ,haTZ)] for the "S2" sandwich configuration of the pyridine dimer at various intermolecular distances (in Å).

Complex	**Original**[a]	**NBC10A**	$\mid \Delta \mid$
PyPy_S2-3.1	2.442	2.387	0.055
PyPy_S2-3.3	-1.125	-1.165	0.040
PyPy_S2-3.4	-2.016	-2.050	0.034
PyPy_S2-3.5	-2.534	-2.562	0.028
PyPy_S2-3.6	-2.791	-2.815	0.024
PyPy_S2-3.7	-2.870	-2.890	0.020
PyPy_S2-3.8	-2.832	-2.849	0.017
PyPy_S2-3.9	-2.719	-2.733	0.014
PyPy_S2-4.0	-2.561	-2.573	0.012
PyPy_S2-4.1	-2.381	-2.391	0.010
PyPy_S2-4.2	-2.192	-2.201	0.009
PyPy_S2-4.3	-2.005	-2.012	0.007
PyPy_S2-4.4	-1.824	-1.830	0.006
PyPy_S2-4.5	-1.655	-1.660	0.005
PyPy_S2-4.7	-1.354	-1.357	0.003
PyPy_S2-5.0	-0.999	-1.002	0.003
PyPy_S2-5.5	-0.618	-0.619	0.001
PyPy_S2-6.0	-0.402	-0.402	0.000
PyPy_S2-6.5	-0.277	-0.276	0.001
PyPy_S2-7.0	-0.200	-0.200	0.000

Table 21: NBC10A benchmark interaction energies [using HF/aQZ, MP2/ CBS(aTZ,aQZ) extrapolation, and $\delta_{MP2}^{CCSD(T)}$/aDZ components] and differences (kcal mol^{-1}) with respect to original values for the "T3" T-shaped configuration of the pyridine dimer at various intermolecular distances (in Å). Unlike the other NBC10 test cases, for T3 pyridine dimer we could not afford CCSD(T)/(h)aTZ computations, hence the level of theory is unchanged. In our revision we reproduced the original values (within a tolerated roundoff error of ± 0.001 kcal mol^{-1}).

Complex	**Original**[a]	**NBC10A**	| Δ |
PyPy_T3-4.1	9.340	9.341	0.001
PyPy_T3-4.3	1.991	1.991	0.000
PyPy_T3-4.5	-1.377	-1.377	0.000
PyPy_T3-4.6	-2.203	-2.203	0.000
PyPy_T3-4.7	-2.673	-2.673	0.000
PyPy_T3-4.8	-2.897	-2.896	0.001
PyPy_T3-4.9	-2.954	-2.954	0.000
PyPy_T3-5.0	-2.903	-2.903	0.000
PyPy_T3-5.1	-2.784	-2.783	0.001
PyPy_T3-5.2	-2.625	-2.625	0.000
PyPy_T3-5.3	-2.447	-2.447	0.000
PyPy_T3-5.4	-2.263	-2.262	0.001
PyPy_T3-5.5	-2.080	-2.080	0.000
PyPy_T3-5.7	-1.742	-1.741	0.001
PyPy_T3-6.0	-1.324	-1.323	0.001
PyPy_T3-6.5	-0.853	-0.852	0.001
PyPy_T3-7.0	-0.574	-0.573	0.001
PyPy_T3-8.0	-0.296	-0.296	0.000
PyPy_T3-9.0	-0.175	-0.174	0.001

REFERENCES

[1] J. Řezác, K. E. Riley, and P. Hobza, J. Chem. Theory Comput., in press.

[2] Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. *MOLPRO*, version 2010.1, see http://www.molpro.net.

[3] Adamo, C., Scuseria, G. E., and Barone, V. *J. Chem. Phys.*, vol. 111, p. 2889, 1999.

[4] Adler, T. B., Knizia, G., and Wener, H.-J. *J. Chem. Phys.*, vol. 127, p. 221106, 2007.

[5] Alvarez-Idaboy, J. R. and Galano, A., "Counterpoise corrected interaction energies are not systematically better than uncorrected ones: Comparison with ccsd(t) cbs extrapolated values," *Theor. Chem. Acc.*, vol. 126, pp. 75–85, 2010.

[6] Amicangelo, J. C. and Armentrout, P. B. *J. Phys. Chem. A*, vol. 104, p. 11420, 2000.

[7] Amunugama, R. and Rodgers, M. T., "Influence of substituents on cation-π interactions. 1. Absolute binding energies of alkali metal cation-toluene complexes determined by threshold collision-induced dissociation and theoretical studies," *J. Phys. Chem. A*, vol. 106, no. 22, pp. 5529–5539, 2002.

[8] Amunugama, R. and Rodgers, M. T., "Influence of substituents on cation-π interactions. 4. Alkali metal cation-phenol complexes determined by threshold collision-induced dissociation and theoretical studies," *J. Phys. Chem. A*, vol. 106, no. 42, pp. 9718–9728, 2002.

[9] Antony, J. and Grimme, S., "Is spin-component scaled second-order møller plesset perturbation theory an appropriate method for the study of noncovalent interactions in molecules?," *J. Phys. Chem. A*, vol. 111, pp. 4862–4868, 2007.

[10] Aquilante, F., Gagliardi, L., Pedersen, T. B., and Lindh, R., "Atomic cholesky decompositions: A route to unbiased auxiliary basis sets for density

fitting approximation with tunable accuracy and efficiency," *J. Chem. Phys.*, vol. 130, p. 154107, 2009.

[11] Aquilante, F. and Pedersen, T. B. *Chem. Phys. Lett.*, vol. 449, p. 354, 2007.

[12] Aquilante, F., Pedersen, T. B., and Lindh, R., "Low-cost evaluation of the exchange fock matrix from cholesky and density fitting representations of the electron repulsion integrals," *J. Chem. Phys.*, vol. 126, p. 194106, 2007.

[13] Armentrout, P. B. and Rodgers, M. T. *J. Phys. Chem. A*, vol. 104, pp. 2238–2247, 2000.

[14] Arunan, E., Desiraju, G. R., Klein, R. A., Sadlej, J., Scheiner, S., Alkorta, I., Clary, D. C., Crabtree, R. H., Dannenberg, J. J., Hobza, P., Kjaergaard, H. G., Legon, A. C., Mennucci, B., and Nesbitt, D. J., "Definition of the hydrogen bond (IUPAC recommendations 2011)," *Pure Appl. Chem.*, vol. 83, pp. 1637–1641, 2011.

[15] Bartlett, R. J., "Many-body perturbation theory and coupled cluster theory for electron correlation in molecules," *Ann. Rev. Phys. Chem.*, vol. 32, pp. 359–401, 1981.

[16] Bartlett, R. J., Sekino, H., and Purvis, G. D., "Comparison of MBPT and coupled-cluster methods with full CI. importance of triplet excitations and infinite summations," *Chem. Phys. Lett.*, vol. 98, pp. 66–71, 1983.

[17] Bartlett, R. J. and Stanton, J. F., "Applications of post-hartree-fock methods: A tutorial," in *Reviews in Computational Chemistry* (Lipkowitz, K. B. and Boyd, D. B., eds.), vol. 5, pp. 65–169, New York: VCH Publishers, 1994.

[18] Becke, A. D. *J. Chem. Phys.*, vol. 98, pp. 5642–5648, 1993.

[19] Beebe, N. H. F. and Linderberg, J., "Simplifications in the generation and transformation of two-electron integrals in molecular calculations," *Int. J. Quantum Chem.*, vol. 12, pp. 683–705, 1977.

[20] Beene, D. L., Brandt, G. S., Zhong, W., Zacharias, N. M., Lester, H. A., and Dougherty, D. A. *Biochemistry*, vol. 41, pp. 10262–10269, 2002.

[21] Bernholdt, D. E. and Harrison, R. J., "Large-scale correlated electronic structure calculations: The ri-mp2 method on parallel computers," *Chem. Phys. Lett.*, vol. 250, pp. 477–484, 1996.

[22] Blaudeau, J. P., McGrath, M. P., Curtiss, L. A., and Radom, L. *J. Chem. Phys.*, vol. 107, p. 5016, 1997.

[23] Boese, A. D., Martin, J. M. L., and Klopper, W., "Basis set limit coupled cluster study of h-bonded systems and assessment of more approximate methods," *J. Phys. Chem. A*, vol. 111, pp. 11122–11133, 2007.

[24] Boström, J., Aquilante, F., Pedersen, T. B., and Lindh, R., "Ab initio density fitting: Accuracy assessment of auxiliary basis sets from cholesky decompositions," *J. Chem. Theory Comput.*, vol. 5, pp. 1545–1553, 2009.

[25] Boström, J., Delcey, M. G., Aquilante, F., Serrano-Andrés, L., Pedersen, T. B., and Lindh, R., "Calibration of cholesky auxiliary basis sets for multiconfigurational perturbation theory calculations of excitation energies," *J. Chem. Theory Comput.*, vol. 6, pp. 747–754, 2010.

[26] Boys, S. F. and Bernardi, F., "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors," *Mol. Phys.*, vol. 19, no. 4, pp. 553–566, 1970.

[27] Brocchieri, L. and Karlin, S. *Proc. Natl. Acad. Sci.*, vol. 91, pp. 9297–9301, 1994.

[28] Burant, J. C., Scuseria, G. E., and Frisch, M. J., "A linear scaling method for hartree-fock exchange calculations of large molecules," *J. Chem. Phys.*, vol. 105, pp. 8969–8972, 1996.

[29] Burns, L. A., Vázquez-Mayagoitia, Á., Sumpter, B. G., and Sherrill, C. D., "Density-functional approaches to noncovalent interactions: A comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals," *J. Chem. Phys.*, vol. 134, p. 084107, 2011.

[30] Cammi, R. and Tomasi, J. *J. Comput. Chem.*, vol. 16, pp. 1449–1458, 1995.

[31] Cheng, Y.-H., Liu, L., Fu, Y., Chen, R., Li, X.-S., and Guo, Q.-X., "Counterion effects on the cation-π interaction between alkaline earth cations and benzene," *J. Phys. Chem. A*, vol. 106, no. 43, pp. 11215–11220, 2002.

[32] Chipot, C., Jaffe, R., Maigret, B., Pearlman, D. A., and Kollman, P. A., "Benzene dimer: A good model for π-π interactions in proteins? a comparison between the benzene and the toluene dimers in the gas phase and in an aqueous solution," *J. Am. Chem. Soc.*, vol. 118, pp. 11217–11224, 1996.

[33] Chwee, T. S. and Carter, E. A., "Cholesky decomposition within local multireference singles and doubles configuration interaction," *J. Chem. Phys.*, vol. 132, p. 074104, 2010.

[34] Clements, A. and Lewis, M. *J. Phys. Chem. A*, vol. 110, pp. 12705–12710, 2006.

[35] Cohen, E. R. and Taylor, B. N. *J. Phys. Chem. Ref. Data*, vol. 17, p. 1795, 1988.

[36] Crowley, P. B. and Golovin, A. *Proteins*, vol. 59, pp. 231–239, 2005.

[37] Császár, A. G., Allen, W. D., and Schaefer, H. F., "In pursuit of the *ab initio* limit for conformational energy prototypes," *J. Chem. Phys.*, vol. 108, no. 23, pp. 9751–9764, 1998.

[38] Csazar, A., Allen, W., and Schaefer, H. *J. Chem. Phys.*, vol. 108, p. 9751, 1998.

[39] Dabkowska, I., Gonzalez, H. V., Jurecka, P., and Hobza, P. *J. Phys. Chem. A*, vol. 109, p. 1131, 2005.

[40] de Lange, K. and Lane, J. *J. Chem. Phys.*, vol. 134, p. 034301, 2011.

[41] Deng, J., Gilbert, A. T. B., and Gill, P. M. W., "Approaching the hartree-fock limit by perturbative methods," *J. Chem. Phys.*, vol. 130, p. 231101, 2009.

[42] Distasio, R., Steele, R., and Head-Gordon, M. *Mol. Phys.*, vol. 105, pp. 2731–2742, 2007.

[43] Doser, B., Lambrecht, D. S., Kussmann, J., and Ochsenfeld, C., "Linear-scaling atomic orbital-based second-order moller-plesset perturbation theory by rigorous integral screening criteria," *J. Chem. Phys.*, vol. 130, p. 064107, 2009.

[44] Dougherty, D. A. *Science*, vol. 271, p. 163, 1996.

[45] Dunbar, R. C., "Metal cation binding to phenol: DFT comparison of the competing sites," *J. Phys. Chem. A*, vol. 106, no. 32, pp. 7328–7337, 2002.

[46] Dunlap, B. I., Connolly, J. W. D., and Sabin, J. R., "Applicability of LCAO-X-alpha methods to molecules containing transition-metal atoms - nickel atom and nickel hydride," *Int. J. Quantum Chem. Symp.*, vol. 11, p. 81, 1977.

[47] Dunlap, B. I., Connolly, J. W. D., and Sabin, J. R., "On some approximations in applications of $X\alpha$ theory," *J. Chem. Phys.*, vol. 71, pp. 3396–3402, 1979.

[48] East, A. and Allen, W. *J. Chem. Phys.*, vol. 99, p. 4638, 1993.

[49] East, A. L. L. and Allen, W. D., "The heat of formation of nco," *J. Chem. Phys.*, vol. 99, no. 6, pp. 4638–4650, 1993.

[50] Faver, J., Benson, M., He, X., Roberts, B., Wang, B., Marshall, M. S., Kennedy, M. R., Sherrill, C. D., and Merz, K. *J. Chem. Theory Comput.*, vol. 7, p. 790, 2011.

[51] Faver, J. C., Benson, M. L., He, X., Roberts, B. P., Wang, B., Marshall, M. S., Kennedy, M. R., Sherrill, C. D., and Merz, K. M., "Formal estimation of errors in computed absolute interaction energies of protein-ligand complexes," *J. Chem. Theory Comput.*, vol. 7, pp. 790–797, 2011.

[52] Feller, D., "A complete basis set estimate of cation-π bond strengths: Na^+(ethylene) and Na^+(benzene)," *Chem. Phys. Lett.*, vol. 322, pp. 543–548, 2000.

[53] Feller, D., Dixon, D. A., and Nicholas, J. B. *J. Phys. Chem. A*, vol. 104, pp. 11414–11419, 2000.

[54] Feyereisen, M., Fitzgerald, G., and Komornicki, A., "Use of approximate integrals in ab initio theory. an application in mp2 calculations," *Chem. Phys. Lett.*, vol. 208, pp. 359–363, 1993.

[55] Fliegl, H., Klopper, W., and Hattig, C., "Coupled-cluster theory with simplified linear-r_{12} corrections: The ccsd(r12) model," *J. Chem. Phys.*, vol. 122, p. 084107, 2005.

[56] Friesner, R. A., Murphy, R. B., Beachy, M. D., Ringnalda, M. N., Pollard, W. T., Dunietz, B. D., and Cao, Y., "Correlated ab initio electronic structure calculations for large molecules," *J. Phys. Chem. A*, vol. 103, pp. 1913–1928, 1999.

[57] Gallivan, J. P. and Dougherty, D. A. *Proc. Natl. Acad. Sci.*, vol. 96, pp. 9459–9464, 1999.

[58] Gauss, J. and Stanton, J. F., "The equilibrium structure of benzene," *J. Phys. Chem. A*, vol. 104, pp. 2865–2868, 2000.

[59] Gerenkamp, M. and Grimme, S., "Scs-mp2 for the calcualation of molecular geometries and harmonic vibrational frequencies," *Chem. Phys. Lett.*, vol. 392, pp. 229–235, 2004.

[60] Gräfenstein, J. and Cremer, D., "An efficient algorithm for the density-functional theory treatment of dispersion interactions," *J. Chem. Phys.*, vol. 130, p. 124105, 2009.

[61] Gráfová, L., Pitoňák, M., Řezáč, J., and Hobza, P., "Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended s22 data set," *J. Chem. Theory Comput.*, vol. 6, pp. 2365–2376, 2010.

[62] Grimme, S., "Improved second-order møller-plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies," *J. Chem. Phys.*, vol. 118, no. 20, pp. 9095–9102, 2003.

[63] Grimme, S., "Accurate description of van der waals complexes by density functional theory including empirical corrections," *J. Comput. Chem.*, vol. 25, pp. 1463–1473, 2004.

[64] Grimme, S., "Semiempirical GGA-type density functional constructed with a long-range dispersion correction," *J. Comput. Chem.*, vol. 27, no. 15, pp. 1787–1799, 2006.

[65] Gromiha, M. M., Santhosh, C., and Ahmad, S. *Biological Macromolecules*, vol. 34, pp. 203–211, 2004.

[66] Grüning, M., Gritsenko, O. V., van Gisbergen, S. J. A., and Baerends, E. J. *J. Chem. Phys.*, vol. 114, p. 652, 2001.

[67] Halkier, A., Helgaker, T., Jørgensen, P., Klopper, W., Koch, H., Olsen, J., and Wilson, A. K., "Basis-set convergence in correlated calculations on Ne, N_2, and H_2O," *Chem. Phys. Lett.*, vol. 286, pp. 243–252, 1998.

[68] Halkier, A., Klopper, W., Helgaker, T., Jørgensen, P., and Taylor, P. R., "Basis set convergence of the interaction energy of hydrogen-bonded complexes," *J. Chem. Phys.*, vol. 111, pp. 9157–9167, 1999.

[69] Halkier, A., Koch, H., Jorgensen, P., Christiansen, O., Nielsen, I. M. B., and Helgaker, T., "A systematic ab initio study of the water dimer in hierarchies of basis sets and correlation models," *Theor. Chem. Acc.*, vol. 97, pp. 150–157, 1997.

[70] Hättig, C., "Optimization of auxiliary basis sets for ri-mp2 and ri-cc2 calculations: Core-valence and quintuple-zeta basis sets for h to ar and qzvpp basis sets for li to kr," *Phys. Chem. Chem. Phys.*, vol. 7, pp. 59–66, 2005.

[71] Hay, B. P. and Bryantsev, V. S. *Chem. Comm.*, vol. 21, pp. 2393–2484, 2008.

[72] Heginbotham, L. and MacKinnon, R. *Neuron*, vol. 8, pp. 483–491, 1992.

[73] Hesselmann, A. and Jansen, G. *J. Phys. Lett.*, vol. 357, p. 464, 2002.

[74] Hesselmann, A. and Jansen, G. *J. Phys. Lett.*, vol. 362, p. 319, 2002.

[75] Hesselmann, A. and Jansen, G. *Phys. Chem. Chem. Phys.*, vol. 122, p. 5010, 2003.

[76] Hesselmann, A. and Jansen, G. *J. Phys. Lett.*, vol. 367, p. 778, 2003.

[77] Hesselmann, A., Jansen, G., and Schütz, M., "Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies," *J. Chem. Phys.*, vol. 122, p. 014103, 2005.

[78] Hill, J. G., Peterson, K. A., Knizia, G., and Werner, H., "Extrapolating mp2 and ccsd explicitly correlated correlation energies to the complete basis set limit with first and second row correlation consistent basis sets," *J. Chem. Phys.*, vol. 131, p. 194105, 2009.

[79] Hill, J. G. and Platts, J. A., "Spin-component scaling methods for weak and stacking interactions," *J. Chem. Theory Comput.*, vol. 3, pp. 80–85, 2007.

[80] Hill, J. G. and Platts, J. A., "Local electron correlation descriptions of the intermolecular stacking interactions between aromatic intercalators and nucleic acids," *Chem. Phys. Lett.*, vol. 479, pp. 279–283, 2009.

[81] Hobza, P., Carsky, P., and Zahradnik, R. *J. Mol. Struct.*, vol. 138, p. 387, 1986.

[82] Hobza, P. and Šponer, J., "Toward true DNA base-stacking energies: MP2, CCSD(T), and complete basis set calculations," *J. Am. Chem. Soc.*, vol. 124, pp. 11802–11808, 2002.

[83] Hobza, P., Sponer, J., and Polasek, M. *J. Am. Chem. Soc.*, vol. 117, p. 792, 1995.

[84] Hobza, P. and Zahradnik, R. *Chem. Rev.*, vol. 88, p. 871, 1988.

[85] Hohenstein, E. G. and Sherrill, C. D., "Wires comput. mol. sci., in press."

[86] Hohenstein, E. G. and Sherrill, C. D. *J. Phys. Chem. A*, vol. 113, pp. 878–886, 2009.

[87] Hohenstein, E. G. and Sherrill, C. D., "Effects of heteroatoms on aromatic π-π interactions: Benzene-pyridine and pyridine dimer," *J. Phys. Chem. A*, vol. 113, pp. 878–886, 2009.

[88] Hohenstein, E. G. and Sherrill, C. D., "Density fitting and Cholesky decomposition approximations in symmetry-adapted perturbation theory: Implementation and application to probe the nature of π-π interactions in linear acenes," *J. Chem. Phys.*, vol. 132, p. 184111, 2010.

[89] Hohenstein, E. G. and Sherrill, C. D., "Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory," *J. Chem. Phys.*, vol. 133, p. 014101, 2010.

[90] Hohenstein, E. G. and Sherrill, C. D., "Efficient evaluation of triple excitations in symmetry-adapted perturbation theory via second-order møller-plesset perturbation theory natural orbitals," *J. Chem. Phys.*, vol. 133, p. 104107, 2010.

[91] Hohenstein, E. G. and Sherrill, C. D., "Wavefunction methods for noncovalent interactions," *WIREs Comput. Mol. Sci.*, vol. 2, pp. 304–326, 2012.

[92] Hopkins, B. W., ElSohly, A. M., and Tschumper, G. S., "Reliable structures and energetics for two new delocalized π-π prototypes: Cyanogen dimer and diacetylene," *Phys. Chem. Chem. Phys.*, vol. 9, pp. 1550–1558, 2007.

[93] Hopkins, B. W. and Tschumper, G. S., "Ab initio studies of $\pi \cdots \pi$ interactions: The effects of quadruple excitations," *J. Phys. Chem. A*, vol. 108, no. 15, pp. 2941–2948, 2004.

[94] Hylleraas, E. A., "A new calculation of the energy of helium in the ground state as well as the lowest term of ortho-helium," *Z. Phys.*, vol. 54, p. 347, 1929.

[95] Janowski, T. and Pulay, P., "High accuracy benchmark calculations on the benzene dimer potential energy surface," *Chem. Phys. Lett.*, vol. 447, pp. 27–32, 2007.

[96] Jansen, G. and Hesselmann, A. *J. Chem. Phys. A*, vol. 105, p. 646, 2001.

[97] Jansen, G. and Hesselmann, A., "Comment on using kohnsham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions," *J. Phys. Chem. A*, vol. 105, pp. 11156–11157, 2001.

[98] Jeziorski, B., Moszynski, R., and Szalewicz, K., "Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes," *Chem. Rev.*, vol. 94, pp. 1887–1930, 1994.

[99] Jung, Y., Lochan, R. C., Dutoi, A. D., and Head-Gordon, M., "Scaled opposite-spin second order Møller-Plesset correlation energy: An economical electronic structure method," *J. Chem. Phys.*, vol. 121, no. 20, pp. 9793–9802, 2004.

[100] Jurecka, P. and Hobza, P. *Chem. Phys. Lett.*, vol. 365, p. 89, 2002.

[101] Jurečka, P. and Hobza, P., "On the convergence of the ($\Delta E^{\mathrm{CCSD(T)}}$ - ΔE^{MP2}) term for complexes with multiple H-bonds," *Chem. Phys. Lett.*, vol. 365, pp. 89–94, 2002.

[102] Jurečka, P. and Hobza, P., "True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine-cytosine, adenine-thymine, and their 9- and 1-methyl derivatives: Complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment," *J. Am. Chem. Soc.*, vol. 125, pp. 15608–15613, 2003.

[103] Jurečka, P., Šponer, J., Černý, J., and Hobza, P., "Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs," *Phys. Chem. Chem. Phys.*, vol. 8, pp. 1985–1993, 2006.

[104] Jurečka, P., Šponer, J., and Hobza, P., "Potential energy surface of the cytosine dimer: MP2 complete basis set limit interaction energies, CCSD(T) correction term, and comparison with the AMBER force field," *J. Phys. Chem. B*, vol. 108, pp. 5466–5471, 2004.

[105] Jurgens-Lutovsky, R. and Almlöf, J., "Dual basis-sets in calculations of electron correlation," *Chem. Phys. Lett.*, vol. 178, pp. 451–454, 1991.

[106] Katz, C. E., Ribelin, T., Withrow, D., Basseri, Y., Manukyan, A. K., Bermudez, A., Buera, C. G., Day, V. W., Powell, D. R., Poutsma, J. L., and Aube, J. *J. Org. Chem.*, vol. 73, pp. 3318–3327, 2008.

[107] Kearney, P. C., Mizoue, L. S., Kumpf, R. A., Forman, J. E., McCurdy, A., and Dougherty, D. A. *J. Am. Chem. Soc.*, vol. 115, pp. 9907–9919, 1993.

[108] Kendall, R. A. and Fruchtl, H. A., "The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development," *Theor. Chem. Acc.*, vol. 97, pp. 158–163, 1997.

[109] Kim, J. and Kim, K. S. *J. Chem. Phys.*, vol. 109, p. 5886, 1998.

[110] Kim, J. S., Lee, S., Cho, S. J., Mhin, B. J., and Kim, K. S., "Structures, energetics, and spectra of aqua-sodium(I): Thermodynamic effects and nonadditive interactions," *J. Chem. Phys.*, vol. 102, pp. 839–849, 1995.

[111] Kim, K. S., Mhin, B. J., Choi, U.-S., and Lee, K., "Ab initio studies of the water dimer using large basis sets: The structure and thermodynamic energies," *J. Chem. Phys.*, vol. 97, pp. 6649–6662, 1992.

[112] Kim, K. S., Tarakeshwar, P., and Lee, J. Y., "Molecular clusters of π-Systems: Theoretical studies of structures, spectra, and origin of interaction energies," *Chem. Rev.*, vol. 100, no. 11, pp. 4145–4185, 2000.

[113] Klopper, W., Manby, F. R., Ten-no, S., and Valeev, E. F., "R12 methods in explicitly correlated molecular electronic structure theory," *Int. Rev. Phys. Chem.*, vol. 25, pp. 427–468, 2006.

[114] Klopper, W., Ruscic, B., Tew, D. P., Bischoff, F. A., and Wolfsegger, S. *Chem. Phys.*, vol. 356, p. 14, 2009.

[115] Klopper, W. and Samson, C. C. M., "Explicitly correlated second-order Møller-Plesset methods with auxiliary basis sets," *J. Chem. Phys.*, vol. 116, pp. 6397–6410, 2002.

[116] Knizia, G., Adler, T., and Werner, H. *J. Chem. Phys.*, vol. 130, p. 054104, 2009.

[117] Koch, H., de Meras, A. S., and Pedersen, T. B., "Reduced scaling in electronic structure calculations using cholesky decompositions," *J. Chem. Phys.*, vol. 118, pp. 9481–9484, 2003.

[118] Kolář, M., Kubař, T., and Hobza, P., "Sequence-dependent configurational entropy change of dna upon intercalation," *J. Phys. Chem. B*, vol. 114, pp. 13446–13454, 2010.

[119] Kolmann, S., Chan, B., and Jordan, M. *Chem. Phys. Lett.*, vol. 467, pp. 126–130, 2008.

[120] Krishnan, R., Binkley, J. S., Seeger, R., and Pople, J. A. *J. Chem. Phys.*, vol. 72, p. 650, 1980.

[121] Kroon-Batenburg, L. M. J. and van Duijneveldt, F. B. *J. Mol. Struct*, vol. 121, p. 185, 1985.

[122] Kutzelnigg, W. and Klopper, W., "Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. I. general theory," *J. Chem. Phys.*, vol. 94, no. 3, pp. 1985–2001, 1991.

[123] Lee, E. C., Kim, D., Jurečka, P., Tarakeshwar, P., Hobza, P., and Kim, K. S., "Understanding of assembly phenomena by aromatic-aromatic interacitons: Benzene dimer and the substituted systems," *J. Phys. Chem. A*, vol. 111, pp. 3446–3457, 2007.

[124] Lee, T. J. and Scuseria, G. E., "Achieving chemical accuracy with coupled-cluster theory," in *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy* (Langhoff, S. R., ed.), pp. 47–108, Dordrecht: Kluwer Academic Publishers, 1995.

[125] Leininger, M. L., Allen, W. D., Schaefer, H. F., and Sherrill, C. D., "Is Møller-plesset perturbation theory a convergent *ab initio* method?," *J. Chem. Phys.*, vol. 112, pp. 9213–9222, 2000.

[126] Li, S., Cooper, V. R., Thonhauser, T., Lundqvist, B. I., and Langreth, D. C., "Stacking interactions and dna intercalation," *J. Phys. Chem. B*, vol. 113, pp. 11166–11172, 2009.

[127] Liang, W. Z. and Head-Gordon, M., "Approaching the basis set limit in density functional theory calculations using dual basis sets without diagonalization," *J. Phys. Chem. A*, vol. 108, pp. 3206–3210, 2004.

[128] Lias, S. G. *J. Phys. Chem. Ref. Data*, vol. 17, 1988.

[129] Ma, J. C. and Dougherty, D. A. *Chem. Rev.*, vol. 97, pp. 1303–1324, 1997.

[130] Mackie, I. D. and DiLabio, G. A., "Approximations to complete basis set-extrapolated, highly correlated non-covalent interaction energies," *J. Chem. Phys.*, vol. 135, p. 134318, 2011.

[131] Manby, F., Werner, H.-J., Adler, T., and May, A. *J. Chem. Phys.*, vol. 124, p. 094103, 2006.

[132] Marchetti, O. and Werner, H., "Accurate calculations of intermolecular interaction energies using explicitly correlated coupled cluster wave functions and a dispersion-weighted mp2 method," *J. Phys. Chem. A*, vol. 113, pp. 11580–11585, 2009.

[133] Marchetti, O. and Werner, H.-J., "Accurate calculations of intermolecular interaction energies using explicitly correlated wave functions," *Phys. Chem. Chem. Phys.*, vol. 10, pp. 3400–3409, 2008.

[134] Marshall, M. S., Burns, L. A., and Sherrill, C. D., "Basis set convergence of the coupled-cluster correction, $\delta_{MP2}^{CCSD(T)}$: Best practices for benchmarking non-covalent interactions and the attendant revision of the S22, NBC10, HBC6, and HSG databases," *J. Chem. Phys.*, vol. 135, p. 194102, 2011.

[135] Martin, J. M. L., "Spectroscopic quality ab initio potential curves for CH, NH, OH, and HF. a convergence study," *Chem. Phys. Lett.*, vol. 292, pp. 411–420, 1998.

[136] Martinez, T. J. and Carter, E. A., "Pseudospectral methods applied to the electron correlation problem," in *Modern Electronic Structure Theory* (Yarkony, D. R., ed.), vol. 2 of *Advanced Series in Physical Chemistry*, pp. 1132–1165, Singapore: World Scientific, 1995.

[137] Martinez, T. J., Mehta, A., and Carter, E. A., "Pseudospectral full configuration interaction," *J. Chem. Phys.*, vol. 97, pp. 1876–1880, 1992.

[138] May, A. J., Valeev, E. F., Polly, R., and Manby, F. R. *Phys. Chem. Chem. Phys.*, vol. 7, p. 2710, 2005.

[139] McLean, A. D. and Chandler, G. S., "Contracted gaussian basis sets for molecular calculations. i. second row atoms, z=11-18," *J. Chem. Phys.*, vol. 72, p. 5639, 1980.

[140] Meadows, E. S., Wall, S. L. D., Barbour, L. J., and Gokel, G. W. *J. Am. Chem. Soc.*, vol. 123, pp. 3092–3107, 2001.

[141] Mecozzi, S., West, A. P., and Dougherty, D. A. *J. Am. Chem. Soc.*, vol. 118, pp. 2307–2308, 1996.

[142] Meyer, E. A., Castellano, R. K., and Diederich, F. *Angew. Chem. Int. Ed.*, vol. 42, pp. 1210–1250, 2003.

[143] Meyer, E. A., Castellano, R. K., and Diederich, F., "Interactions with aromatic rings in chemical and biological recognition," *Angew. Chem., Int. Ed. Engl.*, vol. 42, no. 11, pp. 1210–1250, 2003.

[144] Miertus, S., Scrocco, E., and Tomasi, J. *Chem. Phys.*, vol. 55, pp. 117–129, 1981.

[145] Min, S. K., Lee, E. C., Lee, H. M., Kim, D. Y., Kim, D., and Kim, K. S. *J. Comp. Chem.*, vol. 29, p. 1208, 2007.

[146] Minoux, H. and Chipot, C. *J. Am. Chem. Soc.*, vol. 121, pp. 10366–10372, 1999.

[147] Mishra, B. K., Bajpai, V. K., Ramanathan, V., Gadre, S. R., and Sathyamurthy, N. *Mol. Phys.*, vol. 106, pp. 1557–1566, 2008.

[148] Mu, T. W., Lester, H. A., and Dougherty, D. A. *J. Am. Chem. Soc.*, vol. 125, pp. 6850–6851, 2003.

[149] Nakamura, H., Sakamoto, T., and Wada, A. *Protein Eng.*, vol. 2, p. 177, 1988.

[150] Nelson, D. D., Fraser, G. T., and Klemperer, W., "Does ammonia hydrogen bond?," *Science*, vol. 238, pp. 1670–1674, 1987.

[151] Noga, J., Klopper, W., and Kutzelnigg, W., "CC-R12: An explicitly correlated coupled-cluster theory," in *Recent Advances in Coupled-Cluster Methods* (Bartlett, R. J., ed.), vol. 3 of *Recent Advances in Computational Chemistry*, pp. 1–48, Singapore: World Scientific, 1997.

[152] Noga, J. and Kutzelnigg, W., "Coupled-cluster theory that takes care of the correlation cusp by inclusion of linear terms in the interelectronic coordinates," *J. Chem. Phys.*, vol. 101, pp. 7738–7762, 1994.

[153] Noga, J., Kutzelnigg, W., and Klopper, W., "CC-R12, a correlation cusp corrected coupled-cluster method with a pilot application to the Be_2 potential curve," *Chem. Phys. Lett.*, vol. 199, pp. 497–504, 1992.

[154] Ochsenfeld, C., White, C. A., and Head-Gordon, M., "Linear and sublinear scaling formation of hartree-fock-type exchange matrices," *J. Chem. Phys.*, vol. 109, pp. 1663–1669, 1998.

[155] Pitoňák, M., Janowski, T., Neogrády, P., Pulay, P., and Hobza, P., "Convergence of the ccsd(t) correction term for the stacked complex methyl adenine-methyl thymine: Comparison with lower-cost alternatives," *J. Chem. Theory Comput.*, vol. 5, pp. 1761–1766, 2009.

[156] Pitonak, M., Neogrady, P., and Hobza, P. *Phys. Chem. Chem. Phys.*, vol. 12, p. 1369, 2010.

[157] Pitoňák, M., Neogrády, P., Řezáč, J., Jurečka, P., Urban, M., and Hobza, P., "Benzene dimer: High-level wave function and density functional theory calculations," *J. Chem. Theory Comput.*, vol. 4, pp. 1829–1834, 2008.

[158] Pitoňák, M., Riley, K. E., Neogrády, and Hobza, P., "Highly accurate CCSD(T) and DFT-SAPT stabilization energies of h-bonded and stacked structures of the uracil dimer," *ChemPhysChem*, vol. 9, pp. 1636–1644, 2008.

[159] Pittner, J. and Hobza, P. *Chem. Phys. Lett.*, vol. 390, p. 496, 2004.

[160] Podeszwa, R., Bukowski, R., and Szalewicz, K., "Potential energy surface for the benzene dimer and perturbational analysis of π-π interactions," *J. Phys. Chem. A*, vol. 110, pp. 10345–10354, 2006.

[161] Podeszwa, R., Patkowski, K., and Szalewicz, K., "Improved interaction energy benchmarks for dimers of biological relevance," *Phys. Chem. Chem. Phys.*, vol. 12, pp. 5974–5979, 2010.

[162] Podeszwa, R., Rice, B. M., and Szalewicz, K., "Predicting structure of molecular crystals from first principles," *Phys. Rev. Lett.*, vol. 101, p. 115503, 2008.

[163] Podeszwa, R., Rice, B. M., and Szalewicz, K., "Crystal structure prediction for cyclotrimethylene trinitramine (rdx) from first principles.," *Phys. Chem. Chem. Phys.*, vol. 11, pp. 5512–5518, 2009.

[164] Podeszwa, R. and Szalewicz, K., "Accurate interaction energies for argon, krypton, and benzene dimers from perturbation theory based on the kohn-sham model," *Chem. Phys. Lett.*, vol. 412, pp. 488–493, 2005.

[165] Polyansky, O. L., Császár, A. G., Shirin, S. V., Zobov, N. F., and Barletta, "High-accuracy ab initio rotation-vibration transitions for water," *Science*, vol. 299, pp. 539–542, 2003.

[166] Polyansky, O. L., Zobov, N. F., Viti, S., Tennyson, J., Bernath, P. F., and Wallace, L., "Water on the sun: Line assignments based on variational calculations," *Science*, vol. 277, p. 346, 1997.

[167] Pratuangdejkul, J., Jaudon, P., Ducrocq, C., Nosoongnoen, W., Guerin, G., Conti, M., Loric, S., Launay, J., and Manivet, P. *J. Chem. Theory Comput.*, vol. 2, pp. 746–760, 2006.

[168] Pulay, P. *J. Phys. Lett.*, vol. 73, pp. 323–398, 1980.

[169] Pulay, P. *J. Comp. Chem.*, vol. 3, pp. 556–560, 1982.

[170] Purvis, G. D. and Bartlett, R. J., "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples," *J. Chem. Phys.*, vol. 76, pp. 1910–1918, 1982.

[171] Quinonero, D., Garau, C., Frontera, A., Ballester, P., Costa, A., and Deya, P. *J. Phys. Chem. A*, vol. 109, pp. 4632–4637, 2005.

[172] Raghavachari, K., Trucks, G. W., Pople, J. A., and Head-Gordon, M., "A 5th-order perturbation comparison of electron correlation theories," *Chem. Phys. Lett.*, vol. 157, pp. 479–483, 1989.

[173] Reddy, A. S. and Sastry, G. N. *J. Phys. Chem. A*, vol. 109, pp. 8893–8903, 2005.

[174] Rejnek, J. and Hobza, P. *J. Phys. Chem. B*, vol. 111, p. 641, 2007.

[175] Řezáč, J., Riley, K. E., and Hobza, P., "S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures," *J. Chem. Theory Comput.*, vol. 7, pp. 2427–2438, 2011.

[176] Ringer, A. L., Figgs, M. S., Sinnokrot, M. O., and Sherrill, C. D., "Aliphatic C–H/π interactions: Methane-benzene, methane-phenol, and methane-indole complexes," *J. Phys. Chem. A*, vol. 110, pp. 10822–10828, 2006.

[177] Ringer, A. L., Senenko, A., and Sherrill, C. D., "Models of S/π interactions in protein structures: Comparison of the H_2S-benzene complex with PDB data," *Protein Sci.*, vol. 16, pp. 2216–2223, 2007.

[178] Ringer, A. L. and Sherrill, C. D., "First principles computation of lattice energies of organic solids: The benzene crystal," *Chem. Eur. J.*, vol. 14, pp. 2542–2547, 2008.

[179] Rittby, M. and Bartlett, R. J., "An open-shell spin-restricted coupled cluster method: Application to ionization potentials in N_2," *J. Phys. Chem.*, vol. 92, p. 3033, 1988.

[180] Roeggen, I. and Wisloff-Nilssen, E., "On the beebe-linderberg 2-electron integral approximation," *Chem. Phys. Lett.*, vol. 132, pp. 154–160, 1986.

[181] Romasi, J., Mennucci, B., and Cammi, R. *Chem. Rev.*, vol. 105, pp. 2999–3093, 2005.

[182] Rutledge, L. R., Durst, H. F., and Westmore, S. D. *J. Chem. Theo. Comp.*, vol. 5, p. 1400, 2009.

[183] Rutledge, L. R. and Westmore, S. D. *J. Chem. Theo. Comp.*, vol. 4, p. 1768, 2008.

[184] Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., S., M., Gordon, Jensen, J. J., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Windus, T. L., Dupuis, M., and Montgomery, J. A. *J. Comput. Chem.*, vol. 14, pp. 1347–1363, 1993.

[185] Schutz, M., Brdarski, S., Widmark, P.-O., Lindh, R., and Karlstrom, G. *J. Chem. Phys.*, vol. 107, p. 4597, 1997.

[186] Schwegler, E. and Challacombe, M., "Linear scaling computation of the hartree-fock exchange matrix," *J. Chem. Phys.*, vol. 105, pp. 2726–2734, 1996.

[187] Schwenke, D. W. and Truhlar, D. G. *J. Chem. Phys.*, vol. 82, p. 2418, 1985.

[188] Scuseria, G. E., Janssen, C. L., and Schaefer, H. F., "An efficient reformulation of the closed-shell coupled cluster single and double excitation (CCSD) equations," *J. Chem. Phys.*, vol. 89, p. 7382, 1988.

[189] Scuseria, G. E. and Lee, T. J., "Comparison of coupled-cluster methods which include the effects of connected triple excitations," *J. Chem. Phys.*, vol. 93, p. 5851, 1990.

[190] Scuseria, G. E., Scheiner, A. C., Lee, T. J., Rice, J. E., and Schaefer, H. F., "The closed-shell coupled sluster single and double excitation (CCSD) model for the description of electron correlation. a comparison with configuration interaction (CISD) results," *J. Chem. Phys.*, vol. 86, p. 2881, 1987.

[191] Shao, Y., Molnar, L. F., Jung, Y., Kussmann, J., Ochsenfeld, C., Brown, S. T., Gilbert, A. T. B., Slipchenko, L. V., Levchenko, S. V., O'Neill, D. P., Jr., R. A. D., Lochan, R. C., Wang, T., Beran, G. J. O., Besley, N. A., Herbert, J. M., Lin, C. Y., Voorhis, T. V., Chien, S. H., Sodt, A., Steele, R. P., Rassolov, V. A., Maslen, P. E., Korambath, P. P., Adamson, R. D., Austin, B., Baker, J., Byrd, E. F. C., Dachsel, H., Doerksen, R. J., Dreuw, A., Dunietz, B. D., Dutoi, A. D., Furlani, T. R., Gwaltney, S. R., Heyden, A., Hirata, S., Hsu, C.-P., Kedziora, G., Khalliulin, R. Z., Klunzinger, P., Lee, A. M., Lee, M. S., Liang, W., Lotan, I., Nair, N., Peters, B., Proynov, E. I., Pieniazek, P. A., Rhee, Y. M., Ritchie, J., Rosta, E., Sherrill, C. D., Simmonett, A. C., Subotnik, J. E., Woodcock, H. L., Zhang, W., Bell, A. T., Chakraborty, A. K., Chipman, D. M., Keil, F. J., Warshel, A., Hehre, W. J., Schaefer, H. F., Kong, J., Krylov, A. I., Gill, P. M. W., and Head-Gordon, M., "Advances in methods and algorithms in a modern quantum chemistry program package," *Phys. Chem. Chem. Phys.*, vol. 8, pp. 3172–3191, 2006.

[192] Sherrill, C. D., Sumpter, B. G., Sinnokrot, M. O., Marshall, M. S., Hohenstein, E. G., Walker, R. C., and Gould, I. R., "Assessment of standard force field models against high-quality ab initio potential curves for prototypes of π-π, ch/π, and sh/π interactions," *J. Comput. Chem.*, vol. 30, pp. 2187–2193, 2009.

[193] Sherrill, C. D., Takatani, T., and Hohenstein, E. G., "An assessment of theoretical methods for nonbonded interactions: Comparison to complete basis set limit coupled-cluster potential energy curves for the benzene dimer, the methane dimer, benzene-methane, and benzene-H_2S," *J. Phys. Chem. A*, vol. 113, pp. 10146–10159, 2009.

[194] Shiozaki, T., Kamiya, M., Hirata, S., and Valeev, E. F. *J. Chem. Phys.*, vol. 130, p. 054101, 2009.

[195] Shiozaki, T., Valeev, E. F., and Hirata, S., "Explicitly correlated combined coupled-cluster and perturbation methods," *J. Chem. Phys.*, vol. 131, p. 044118, 2009.

[196] Singh, N. J., Min, S. K., Kim, D. Y., and Kim, K. S. *J. Chem. Theory Comput.*, vol. 5, pp. 515–529, 2009.

[197] Sinnokrot, M. O. and Sherrill, C. D., "Highly accurate coupled cluster potential energy curves for benzene dimer: The sandwich, T-shaped, and parallel-displaced configurations," *J. Phys. Chem. A*, vol. 108, no. 46, pp. 10200–10207, 2004.

[198] Sinnokrot, M. O. and Sherrill, C. D., "High-accuracy quantum mechanical studies of π-π interactions in benzene dimers," *J. Phys. Chem. A*, vol. 110, pp. 10656–10668, 2006.

[199] Sinnokrot, M. O., Valeev, E. F., and Sherrill, C. D., "Estimates of the ab initio limit for π-π interactions: The benzene dimer," *J. Am. Chem. Soc.*, vol. 124, pp. 10887–10893, 2002.

[200] Sodt, A., Beran, G. J. O., Jung, Y. S., Austin, B., and Head-Gordon, M., "A fast implementation of perfect pairing and imperfect pairing using the resolution of the identity approximation," *J. Chem. Theory Comput.*, vol. 2, pp. 300–305, 2006.

[201] Sodt, A., Subotnik, J. E., and Head-Gordon, M. *J. Chem. Phys.*, vol. 125, p. 194109, 2006.

[202] Soteras, I., Orozco, M., and Luque, F. J. *Phys. Chem. Chem. Phys.*, vol. 10, pp. 2616–2624, 2008.

[203] Steele, R. and Head-Gordon, M. *Mol. Phys.*, vol. 105, pp. 2455–2473, 2007.

[204] Steele, R., Shao, Y., DiStasio, R., and Head-Gordon, M. *J. Chem. Phys. A*, vol. 110, pp. 13915–13922, 2006.

[205] Steele, R. P., DiStasio, R. A., and Head-Gordon, M., "Non-covalent interactions with dual-basis methods: Pairings for augmented basis sets," *J. Chem. Theory Comput.*, vol. 5, pp. 1560–1572, 2009.

[206] Steele, R. P., DiStasio, R. A., Shao, Y., Kong, J., and Head-Gordon, M., "Dual-basis second-order moller-plesset perturbation theory: A reduced-cost reference for correlation calculations," *J. Chem. Phys.*, vol. 125, p. 074108, 2006.

[207] Strain, M. C., Scuseria, G. E., and Frisch, M. J., "Achieving linear scaling for the electronic quantum coulomb problem," *Science*, vol. 271, pp. 51–53, 1996.

[208] Sunner, J., Nishizawa, K., and Kerbarle, P. *J. Phys. Chem.*, vol. 85, p. 1814, 1981.

[209] Szabo, A. and Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. New York: McGraw-Hill, 1989.

[210] Szalewicz, K., Jeziorski, B., Monkhorst, H. J., and Zabolitzky, J. H. *J. Chem. Phys.*, vol. 78, p. 1420, 1983.

[211] Takatani, T., Hohenstein, E. G., Malagoli, M., Marshall, M. S., and Sherrill, C. D., "Basis set consistent revision of the S22 test set of noncovalent interaction energies," *J. Chem. Phys.*, vol. 132, p. 144104, 2010.

[212] Takatani, T., Hohenstein, E. G., and Sherrill, C. D., "Improvement of the coupled-cluster singles and doubles method via scaling same- and opposite-spin components of the double excitation correlation energy," *J. Chem. Phys.*, vol. 128, p. 124111, 2008.

[213] Takatani, T. and Sherrill, C. D., "Performance of spin-component-scaled Møller-Plesset theory (SCS-MP2) for potential energy curves of noncovalent interactions," *Phys. Chem. Chem. Phys.*, vol. 9, pp. 6106–6114, 2007.

[214] Temelso, B., Valeev, E. F., and Sherrill, C. D., "A comparison of one-particle basis set completeness, higher-order electron correlation, relativistic effects, and adiabatic corrections for spectroscopic constants of BH, CH^+, and NH," *J. Phys. Chem. A*, vol. 108, pp. 3068–3075, 2004.

[215] Ten-no, S., "Initiation of explicitly correlated Slater-type geminal theory," *Chem. Phys. Lett.*, vol. 398, pp. 56–61, 2004.

[216] Termath, V., Klopper, W., and Kutzelnigg, W., "Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. II. second-order Møller-Plesset (MP2-R12) calculations on closed-shell atoms," *J. Chem. Phys.*, vol. 94, no. 3, pp. 2002–2030, 1991.

[217] Tew, D. P., Klopper, W., Neiss, C., and Hattig, C. *Phys. Chem. Chem. Phys.*, vol. 9, p. 1921, 2007.

[218] Thanthiriwatte, K. S., Hohenstein, E. G., Burns, L. A., and Sherrill, C. D., "Assessment of the performance of dft and dft-d methods for describing distance dependence of hydrogen-bonded interactions," *J. Chem. Theory Comput.*, vol. 7, pp. 88–96, 2011.

[219] Tomasi, J. and Persico, M. *Chem. Rev.*, vol. 94, pp. 2027–2094, 1994.

[220] Torheyden, M. and Valeev, E. F., "Universal perturbative explicitly correlated basis set incompleteness correction," *J. Chem. Phys.*, vol. 131, p. 171103, 2009.

[221] Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M., and Tanabe, K., "Origin of attraction and directionality of the π-π interaction: Model chemistry calculations of benzene dimer interaction," *J. Am. Chem. Soc.*, vol. 124, no. 1, pp. 104–112, 2002.

[222] Tsuzuki, S., Mikami, M., and Yamada, S. *J. Am. Chem. Soc.*, vol. 129, pp. 8656–8662, 2007.

[223] Tsuzuki, S., Yoshida, M., Uchimaru, T., and Mikami, M. *J. Phys. Chem. A*, vol. 105, pp. 769–773, 2001.

[224] Valeev, E. F., "Improving on the resolution of the identity in linear R12 ab initio theories," *Chem. Phys. Lett.*, vol. 395, pp. 190–195, 2004.

[225] Valeev, E. F., "Combining explicitly correlated r12 and gaussian geminal electronic structure theories," *J. Chem. Phys.*, vol. 125, p. 244106, 2006.

[226] Weigend, F., "A fully direct ri-hf algorithm: Implementation, optimized auxiliary basis sets, demonstration of accuracy and efficiency," *Phys. Chem. Chem. Phys.*, vol. 4, pp. 4285–4291, 2002.

[227] Weigend, F., Kattannek, M., and Ahlrichs, R., "Approximated electron repulsion integrals: Cholesky decomposition versus resolution of the identity methods," *J. Chem. Phys.*, vol. 130, p. 164106, 2009.

[228] Weigend, F., Köhn, A., and Hättig, C., "Efficient use of the correlation consistent basis sets in resolution of the identity mp2 calculations," *J. Chem. Phys.*, vol. 116, pp. 3175–3183, 2002.

[229] Werner, H., Adler, T., and Manby, F. *J. Chem. Phys.*, vol. 126, p. 164102, 2007.

[230] Werner, H. and Manby, F., "Explicitly correlated second-order perturbation theory using density fitting and local approximations," *J. Chem. Phys.*, vol. 124, p. 054114, 2006.

[231] Werner, H.-J. *J. Chem. Phys.*, vol. 124, p. 101103, 2008.

[232] Werner, H. J., Knowles, P. J., Lindh, R., Manby, F. R., Schütz, M., Celani, P., Korona, T., Rauhut, G., Amos, R. D., Bernhardsson, A., Berning, A., Cooper, D. L., Deegan, M. J. O., Dobbyn, A. J., Eckert, F., Hampel, C., Hetzer, G., Lloyd, A. W., McNicholas, S. J., Meyer, W., Mura, M. E., Nicklass, A., Palmieri, P., Pitzer, R., Schumann, U., Stoll, H., Stone, A. J., Tarroni, R., and Thorsteinsson, T., "Molpro, version 2006.1, a package of ab initio programs," 2006. see http://www.molpro.net.

[233] Werner, H.-J., Manby, F. R., and Knowles, P. J., "Fast linear scaling second-order møller-plesset perturbation theory (MP2) using local and density fitting approximations," *J. Chem. Phys.*, vol. 118, no. 18, pp. 8149–8160, 2003.

[234] White, C., Johnson, B. G., Gill, P. M. W., and Head-Gordon, M., "Linear scaling density functional calculations via the continuous fast multipole method," *Chem. Phys. Lett.*, vol. 253, pp. 268–278, 1996.

[235] Whitten, J. L., "Coulombic potential-energy integrals and approximations," *J. Chem. Phys.*, vol. 58, pp. 4496–4501, 1973.

[236] Williams, H. L. and Chabalowski, C. F., "Using kohn-sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions," *J. Phys. Chem. A*, vol. 105, pp. 646–659, 2001.

[237] Williams, H. L., Szalewicz, K., Jeziorski, B., Moszynski, R., and Rybak, S., "Symmetry-adapted perturbation theory calculation of the Ar-H_2 intermolecular potential energy surface," *J. Chem. Phys.*, vol. 98, pp. 1279–1292, 1993.

[238] Wolinski, K. and Pulay, P., "Second-order moller-plesset calculations with dual basis sets," *J. Chem. Phys.*, vol. 118, pp. 9497–9503, 2003.

[239] Woolf, T. B., Grossfield, A., and Pearson, J. G. *Int. J. Quantum Chem.*, vol. 75, p. 197, 1999.

[240] Wu, Q. and Yang, W., "Empirical correction to density functional theory for van der waals interactions," *J. Chem. Phys.*, vol. 116, pp. 515–524, 2002.

[241] Wu, R. and McMahon, T. B. *J. Am. Chem. Soc.*, vol. 130, pp. 12554–12555, 2008.

[242] Yuan, X. L., Cheng, X. L., and Su, X. F. *Int. J. Quan. Chem.*, vol. 109, pp. 972–981, 2009.

[243] Zhao, Y. and Truhlar, D. G. *J. Phys. Chem. A*, vol. 109, p. 6624, 2005.

[244] Zhong, W., Gallivan, J. P., Zhang, Y., Li, L., Lester, H. A., and Dougherty, D. A. *Proc. Natl. Acad. Sci.*, vol. 95, p. 12088, 1998.

[245] Zienau, J., Clin, L., Doser, B., and Ochsenfeld, C. *J. Chem. Phys.*, vol. 130, p. 204112, 2009.

[246] Zienau, J., Clin, L., Doser, B., and Ochsenfeld, C., "Cholesky-decomposed densities in laplace-based second-order møller-plesset perturbation theory," *J. Chem. Phys.*, vol. 130, p. 204112, 2009.

VITA

Michael S. Marshall was born in 1984 in Montgomery, Alabama to the parents of Edward M. Marshall and Aldeana J. Marshall. He graduated high-school in 2002 from Holtville high-school in Slapout, Alabama. He then attended the University of Alabama at Birmingham, where he worked with Dr. Tracy Hamilton, graduating *magna cum lauda* with a degree in Physics. He then spent the next five years studying computational chemistry under the guidance of Dr. C. David Sherrill as a Presidential Fellow.